

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:44:37 ; Search time 92,967 Seconds  
(without alignments)  
495.736 Million cell updates/sec

Title: US-09-975-502A-8

Perfect score: 458

Sequence: 1 MKFLAVLVLLGVIFLVSQA.....KDIPVLPKWGDLNGRVCVP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : UniProt 03:\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	458	100.0	90	2 Q96DR8	Q96dr8 homo sapien
2	120.5	26.3	1349	2 Q8WQ04	Q8wq04 homo sapien
3	115	25.1	917	2 Q88U0J	Q88uj0 lactobacill
4	112.5	24.6	208	2 Q62265	Q62265 mus musculu
5	110.5	24.1	182	2 Q9U9J3	Q9u9j3 toxocara ca
6	110	24.0	173	1 HA34_BRELC	Q99074 bremla lact
7	108.5	23.7	164	2 Q14851	Q14851 homo sapien
8	108	23.6	67	2 Q96DU6	Q96duc homo sapien
9	107.5	23.5	135	2 Q9VY79	Q9vyt9 drosophila
10	107	23.4	279	2 Q14888	Q14888 homo sapien
11	107	23.4	788	2 Q18510	Q18510 trichoplusi
12	107	23.4	807	2 Q18511	Q18511 trichoplusi
13	107	23.4	2338	2 Q629M6	Q629nc burkholderi
14	106.5	23.3	71	2 Q6ML46	Q6ml46 bdellovibri
15	106.5	23.3	208	2 Q91X55	Q91x55 mus musculu
16	105.5	23.0	907	1 A180_HUMAN	Q60641 homo sapien
17	105	22.9	184	2 Q84L71	Q84l71 phytophthor
18	105	22.9	565	2 Q73T86	Q73t86 mycobacteri
19	104	22.7	105	2 Q9PEK4	Q9pek4 xyliella fas
20	104	22.7	130	2 Q81Q88	Q81q88 drosophila
21	104	22.7	130	2 Q9U519	Q9u519 drosophila
22	104	22.7	136	2 Q81Q88	Q81q88 drosophila
23	103.5	22.6	2448	2 Q8WQ05	Q8wq05 homo sapien
24	103	22.5	816	2 Q88DU9	Q88dl9 homo sapien
25	102.5	22.4	1787	2 Q25645	Q25645 plasmodium
26	102.5	22.4	1791	2 Q9U6D4	Q9u6d4 plasmodium
27	101.5	22.2	600	1 SP96_DICTDI	P14328 dictyosteli
28	101.5	22.2	600	1 Q86B01	Q86b01 dictyosteli
29	100.5	21.9	811	2 Q95U94	Q95u94 mamestra co
30	100	21.8	191	2 Q9U9J1	Q9u9j1 toxocara ca
31	100	21.8	477	2 Q14887	Q14887 homo sapien

ALIGNMENTS

32	99.5	21.7	88	2	Q7KUG3	Q7kug3 drosophila
33	99.5	21.7	301	2	P74823	P74823 sphingomona
34	99.5	21.7	309	2	Q63549	Q63549 rattus norv
35	99.5	21.7	316	2	Q8U9J0	Q8u9j0 toxocara ca
36	99	21.6	316	2	Q8VQ04	Q8vq04 equid herpe
37	99	21.6	662	1	MUC1_XENLA	Q05049 xenopus lae
38	98.5	21.5	269	2	Q9U9J2	Q9u9j2 toxocara ca
39	98.5	21.5	481	2	Q6CIA5	Q6cia5 yarrowia li
40	98	21.4	88	2	Q8JHE4	Q8jhe4 neotoma
41	98	21.4	135	2	Q00876	Q00876 plasmodium
42	98	21.4	260	2	Q7PR65	Q7pr65 anopheles g
43	97.5	21.3	72	2	Q90401	Q90401 dissoctichu
44	97.5	21.3	285	2	Q9DAT1	Q9dat1 m mus muscu
45	97.5	21.3	347	2	Q8SYW5	Q8syw5 drosophila
46	97.5	21.3	471	2	Q9VMG7	Q9vmg7 drosophila
47	97.5	21.3	608	2	Q8CHE0	Q8che0 mus musculu
48	97.5	21.3	627	2	Q7TT20	Q7tt20 mus musculu
49	97.5	21.3	868	2	Q8K0D4	Q8k0d4 mus musculu
50	97.5	21.3	901	1	A180_MOUSE	Q05140 rattus norv
51	97.5	21.3	915	1	A180_RAT	Q05140 rattus norv
52	97.5	21.3	1030	2	Q962B3	Q962b3 helicoverpa
53	96.5	21.1	88	2	Q7KUG2	Q7kug2 drosophila
54	96.5	21.1	88	2	Q7KUG4	Q7kug4 drosophila
55	96.5	21.1	88	2	Q7KUG8	Q7kug8 drosophila
56	96.5	21.1	88	2	Q7KUG9	Q7kug9 drosophila
57	96.5	21.1	95	2	Q7KUG7	Q7kug7 drosophila
58	96.5	21.1	332	2	Q8V0L9	Q8v0l9 equid herpe
59	96.5	21.1	337	2	Q8V0M1	Q8v0m1 equid herpe
60	96.5	21.1	342	2	Q8V0L8	Q8v0l8 equid herpe
61	96.5	21.1	356	2	Q8V0L7	Q8v0l7 equid herpe
62	96.5	21.1	357	2	Q8V0M2	Q8v0m2 equid herpe
63	96.5	21.1	372	2	Q8V0M3	Q8v0m3 equid herpe
64	96.5	21.1	389	2	Q8V0M0	Q8v0m0 equid herpe
65	96.5	21.1	791	2	Q66VC3	Q66vc3 equid herpe
66	96.5	21.1	797	1	VGLX_EHV1B	P28968 equine herp
67	96.5	21.1	797	2	Q6DL60	Q6dl60 equid herpe
68	96.5	21.1	826	2	Q8V0L5	Q8v0l5 equid herpe
69	96.5	21.1	866	2	Q39781	Q39781 equid herpe
70	96.5	21.1	866	2	Q6S6W0	Q6s6w0 equid herpe
71	96.5	21.1	867	2	Q39782	Q39782 equid herpe
72	96	21.0	87	2	Q6VEL6	Q6vel6 drosophila
73	96	21.0	201	2	Q8V0K4	Q8v0k4 equid herpe
74	96	21.0	217	2	Q8V0K2	Q8v0k2 equid herpe
75	96	21.0	218	2	Q8V0K3	Q8v0k3 equid herpe
76	96	21.0	240	2	Q8V0K5	Q8v0k5 equid herpe
77	96	21.0	245	2	Q8V0K7	Q8v0k7 equid herpe
78	96	21.0	258	2	Q8V0K9	Q8v0k9 equid herpe
79	96	21.0	260	2	Q8V0K6	Q8v0k6 equid herpe
80	96	21.0	374	2	Q8V0L6	Q8v0l6 equid herpe
81	96	21.0	703	2	Q8V0L3	Q8v0l3 equid herpe
82	96	21.0	726	2	Q8V0L4	Q8v0l4 equid herpe
83	96	21.0	750	2	Q39307	Q39307 equid herpe
84	96	21.0	779	2	Q8V0L2	Q8v0l2 equid herpe
85	96	21.0	804	2	Q8V0L1	Q8v0l1 equid herpe
86	96	21.0	825	2	Q8V0L0	Q8v0l0 equid herpe
87	95.5	20.9	207	2	Q26879	Q26879 trypanosoma
88	95.5	20.9	322	2	Q62605	Q62605 rattus norv
89	95.5	20.9	653	2	Q73TB8	Q73tb8 mycobacteri
90	95.5	20.9	722	2	Q13083	Q13083 dissoctichu
91	95.5	20.9	904	2	Q82HW9	Q82hw9 streptomyce
92	95	20.7	135	2	Q00875	Q00875 plasmodium
93	95	20.7	172	2	Q7PRG9	Q7prg9 anopheles g
94	95	20.7	3295	2	Q66GT3	Q66gt3 rattus norv
95	94.5	20.6	481	2	Q9XUF4	Q9xuf4 caenorhabdi
96	94	20.5	94	2	Q14886	Q14886 homo sapien
97	94	20.5	117	2	Q8K0W1	Q8k0w1 mus musculu
98	94	20.5	138	1	SG16_MOUSE	P02815 mus musculu
99	94	20.5	138	2	Q9DA65	Q9da65 mus musculu
100	94	20.5	146	2	Q76915	Q76915 drosophila

RESULT 1  
 Q96DR8 PRELIMINARY; PRT; 90 AA.  
 ID Q96DR8  
 AC Q96DR8;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 05-JUN-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Small breast epithelial mucin (KFLA590).  
 GN ORFNames=UNQ590;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Normal breast tissue and breast tumor;  
 RX MEDLINE=22013742; PubMed=12019145;  
 RA Mksicek R.J., Myal Y., Watson P.H., Walker C., Murphy L.C.,  
 RA Leygue E.;  
 RA "Identification of a novel breast- and salivary gland-specific, mucin-  
 RT like gene strongly expressed in normal and tumor human mammary  
 RT epithelium.";  
 RL Cancer Res. 62:2736-2740 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22887296; PubMed=12975309; DOI=10.1101/gr.1293003;  
 RA Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,  
 RA Chen J., Chow B., Chui C., Crowley C., Currell B., Devel B., Dowd P.,  
 RA Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.E., Heldens S.,  
 RA Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,  
 RA Lewis L., Liao D., Mark M., Robble E., Sanchez C., Schoenfeld J.,  
 RA Seehagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,  
 RA Vandlen R., Watanabe C., Wiewand D., Woods K., Xie M.H., Yansura D.,  
 RA Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,  
 RA Godowski P.;  
 RA "The secreted protein discovery initiative (SPDI), a large-scale  
 RT effort to identify novel human secreted and transmembrane proteins: a  
 RT bioinformatics assessment.";  
 RL Genome Res. 13:2265-2270 (2003).  
 DR ENBL; AF414087; AAL02119.1; -  
 DR EMBL; AY359062; AA089421.1; -  
 SQ SEQUENCE 90 AA; 9039 MW; FEA58A2833B07262 CRC64;  
  
 Query Match 100.0%; Score 458; DB 2; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-32;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 MKFLAVLVLGVSIFLVSQAQNTTAAADTYPATGPADPADTAAATTTAAPTT 60  
 DB 1 MKFLAVLVLGVSIFLVSQAQNTTAAADTYPATGPADPADTAAATTTAAPTT 60  
  
 QY 61 ATTAATTAARKDIPVLPKKGWGDLENGRVCP 90  
 DB 61 ATTAATTAARKDIPVLPKKGWGDLENGRVCP 90  
  
 RESULT 2  
 Q8WQW4 PRELIMINARY; PRT; 1349 AA.  
 ID Q8WQW4  
 AC Q8WQW4;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Mucin 5 (Fragment).  
 GN Name=MUC5AC;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21426417; PubMed=11535137; DOI=10.1042/0264-6021:35980763;

01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
Salivary protein 2 precursor.  
Name=Spt2;  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=89158788; PubMed=2921944;  
RA Dickinson D.P., Mirsels L., Tabak L.A., Gross K.W.;  
RT "Rapid evolution of variants in a rodent multigene family encoding  
salivary proteins".  
RL Mol. Biol. Evol. 6:80-102(1989).  
[2]  
SEQUENCE FROM N.A.  
RX STRAIN=FVB/N; TISSUE=Salivary gland;  
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Collins B., Wagner L., Shenmen C.M., Schuler G.D.,  
Klausner R.D., Collins B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,  
Rosak S.A., Morley P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E.,  
Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[3]  
SEQUENCE FROM N.A.  
RX STRAIN=FVB/N; TISSUE=Salivary gland;  
RC STRAUSBERG R.;  
RA Strausberg R.;  
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE FROM N.A.  
RX STRAIN=FVB/N; TISSUE=Salivary gland;  
RC STRAUSBERG R.;  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; M33975; AAA40133.1; -;  
DR EMBL; BC059095; AAH59095.1; -;  
DR EMBL; BC024641; AAH24641.1; -;  
DR PIR; B30305; B30305.  
DR MGD; MGI:98393; Spt2.  
DR GO; GO:0005615; C:extracellular space; TAS.  
KW Signal.  
FT SIGNAL  
FT CHAIN  
FT SIGNAL  
FT CHAIN  
SQ SEQUENCE 208 AA; 22852 MW; EBIDAB139170823F CRC64;  
Query Match 24.6%; Score 112.5; DB 2; Length 208;  
Best Local Similarity 44.6%; Pred. No. 0.045;  
Matches 33; Conservative 9; Mismatches 23; Indels 9; Gaps 3;  
QY 1 MKFLAVLLVGVISFLVSAQNPPTAAPADTYPATGTPADDEAPDAET--TAAATATTAAAP 58  
DB 1 MKFLAVLLVGVISFLVSAQNPPTAAPADTYPATGTPADDEAPDAET--TAAATATTAAAP 58  
QY 59 TTATTAASTARKD 72  
DB 57 TQAD---STNENQD 67

RESULT 5  
Q9U9J3 PRELIMINARY; PRT; 182 AA.  
AC Q9U9J3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Excretory/secretory mucin MUC-2.  
GN Name=muc-2;  
OS Toxocara canis (Canine roundworm).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;  
OC Toxocaridae; Toxocara.  
OX NCBI\_TaxID=6265;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99386876; PubMed=10456930;  
RA Tetteh K.K., Loukas A., Tripp C., Maizels R.M.;  
RT "Identification of abundantly expressed novel and conserved genes from  
the infective larval stage of Toxocara canis by an expressed sequence  
tag strategy".  
RL Infect. Immun. 67:4771-4779(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Tetteh K.K., Loukas A., Maizels R.M.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF167707; AAD49339.1; -;  
DR InterPro; IPR003582; SHK1.  
DR Pfam; PF01549; SHK1; 2.  
DR SMART; SM00254; SHK1; 2.  
SQ SEQUENCE 182 AA; 18109 MW; 9DDB9A87F1E46DE9 CRC64;  
Query Match 24.1%; Score 110.5; DB 2; Length 182;  
Best Local Similarity 46.5%; Pred. No. 0.059;  
Matches 33; Conservative 4; Mismatches 27; Indels 7; Gaps 2;  
QY 6 VLVLGVISFLV----SAQNPPTAAPADTYPATGTPADDEAPDAETAAATATTAAAP 58  
DB 5 VLLVTVLISVVKPQPGQQTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAP 64  
QY 59 TTATTAASTTA 69  
DB 65 ITTAGATTTA 75  
RESULT 6  
HA34 BRELC STANDARD; PRT; 173 AA.  
AC Q95074;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE HAM34 protein.  
GN Name=HAM34;  
OS Bremia lactucae (Lettuce downy mildew).  
OC Eukaryota; stramenopiles; Oomycetes; Peronosporaceae;  
OC Bremia.  
OX NCBI\_TaxID=4779;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=REGEL; TISSUE=Spore;  
RX MEDLINE=92033077; PubMed=2131094;  
RA Judelson H.S., Michelmore R.W.;  
RT "Highly abundant and stage-specific mRNAs in the obligate pathogen  
Bremia lactucae".  
RL Mol. Plant Microbe Interact. 3:225-232(1990).  
CC -!- FUNCTION: Could be a structural protein required for the infection  
process of B.lactucae.  
CC -!- TISSUE SPECIFICITY: Germinating spores.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way





RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*.";  
RA Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537569;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RA "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*  
RA melanogaster euchromatic genome sequence.";  
RA Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RA "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RA a genomic perspective.";  
RA Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bertram B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RA systematic review.";  
RA Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
RN [5]  
RP SEQUENCE FROM N.A.  
RG FlyBase;  
RN Submitted (SRP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RG FlyBase;  
RN Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003487; AAF48099.1; -;  
DR InTact; OSVT9; -;  
DR FlyBase; FBGN0030338; CG15741.  
SQ SEQUENCE 135 AA; 14154 MW; 2F6E1D563898467A CRC64;  
Query Match 23.5%; Score 107.5; DB 2; Length 135;  
Best Local Similarity 32.3%; Pred. No. 0.081;  
Matches 30; Conservative 15; Mismatches 25; Indels 23; Gaps 4;  
Qy 1 MKFLAVLVLLGVSLFVLSAQNPHTAAPTTPA-----TGPADEAPDAET----- 46  
Db 1 MRFLCVLIL--ASLLAVASSTSPASSTSPASSTSPASSTSPASSTSPASSTSPSS 58  
Qy 47 -----TAAATATTAAAPTATT--AASTARKD 72  
Db 59 SSSSSSTATTATTTTVAPTTTTTSASSSSSSD 91  
RESULT 10

Q14888  
ID Q14888 PRELIMINARY; PRT; 279 AA.  
AC Q14888;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Mucin (Fragment).  
DE Name=MUC5AC;  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=tracheobronchial mucosa;  
RA Guyonnet-Duperat V., Audie J., Debailleul V., Laine A., Buisine M.,  
RA Zouitina-Gallieue S., Pigny P., Degand P., Aubert J., Porchet N.;  
RA "Characterization of the human mucin gene MUC5AC: a consensus  
RA cysteine-rich domain for 11p15 mucin genes.";  
RL Biochem. J. 0:0-0(1994).  
DR EMBL; Z34278; CAA84032.1; -;  
DR PIR; S53363; S53363.  
FT NON\_TER 1 1  
FT NON\_TER 279 279  
SQ SEQUENCE 279 AA; 26254 MW; 0D3465C151C3C25A CRC64;  
Query Match 23.4%; Score 107; DB 2; Length 279;  
Best Local Similarity 44.4%; Pred. No. 0.18;  
Matches 24; Conservative 10; Mismatches 18; Indels 2; Gaps 1;  
Qy 18 SAQNPTTAAPADTYPATGP--ADDEAPDAETAAATTAATTAATTAATTAATTAATTA 69  
Db 34 SAPKSTSAATTTTSGPTTPRPVPTTSTTSPTTSTTSPTTSTTSATTS 87  
RESULT 11  
O18510  
ID O18510 PRELIMINARY; PRT; 788 AA.  
AC O18510;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Insect intestinal mucin IIM14.  
OS Trichoplusia ni (Cabbage looper).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
OC Noctuidae; Plusiinae; Trichoplusia.  
OC NCBI\_TaxID=7111;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97341213; PubMed=9195982; DOI=10.1074/jbc.272.26.16663;  
RA Wang P., Granados R.R.;  
RA "Molecular cloning and sequencing of a novel invertebrate intestinal  
RA mucin cDNA.";  
RL J. Biol. Chem. 272:16663-16669(1997).  
DR EMBL; AF000605; AAC47556.1; -;  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008061; P:chitin binding; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin\_bind\_Pera.  
DR Pfam; PF01607; CBM 14; 5.  
DR SMART; SM00494; ChEBD2; 5.  
DR PROSITE; PS00940; CHIT\_BIND II; 5.  
SQ SEQUENCE 788 AA; 81716 MW; FB7928D9E43B7A19 CRC64;  
Query Match 23.4%; Score 107; DB 2; Length 788;  
Best Local Similarity 39.2%; Pred. No. 0.47;  
Matches 38; Conservative 3; Mismatches 30; Indels 26; Gaps 5;  
Qy 18 SAQNPTTAAPADTYPATGPADDEAP-----DAETAAAT--ATAAPT 59  
Db 157 TQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTT 216

```
QY 60 T--ATTAAGTT-----ARKDIPVLPKWVDLPNGRVCP 90
Db 217 TPAATTAATPGVPAPTSAPVWPPIICELLPNG--CP 251

RESULT 12
Q18511
ID O18511 PRELIMINARY; PRT; 807 AA.
AC O18511;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Insect intestinal mucin IIM22.
OS Trichoplusia ni (Cabbage looper).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Nooptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Plusiinae; Trichoplusia.
OX NCBI_TaxID=7111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97341213; PubMed=9195982; DOI=10.1074/jbc.272.26.16663;
RA Wang P., Granados R.R.;
RT "Molecular cloning and sequencing of a novel invertebrate intestinal
mucin cDNA.";
RL J. Biol. Chem. 272:16663-16669(1997).
DR EMBL; AF000606; AAC47557.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_Pera.
DR Pfam; PF01607; CBM14; 5.
DR SMART; SM00494; ChtBD2; 5.
DR PROSITE; PS00940; CHIT BIND II; 5.
SQ SEQUENCE 807 AA; 83408 MW; 08AF993B0A342892 CRC64;

Query Match 23.4%; Score 107; DB 2; Length 807;
Best Local Similarity 39.2%; Pred. No. 0.48;
Matches 38; Conservative 3; Mismatches 30; Indels 26; Gaps 5;

QY 18 SAQNPTTAADTPATGPDADP-----DAETTAAT--ATTAAPT 59
Db 157 TQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAATTPAATTPAAT 216

QY 60 T--ATTAAGTT-----ARKDIPVLPKWVDLPNGRVCP 90
Db 217 TPAATTAATPGVPAPTSAPVWPPIICELLPNG--CP 251

RESULT 13
Q629N6
ID Q629N6 PRELIMINARY; PRT; 2338 AA.
AC Q629N6;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Polyketide synthase, putative.
GN ORFNames=BMAA2089;
OS Burkholderia mallei ATCC 23344.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=243160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23344;
RA Nierman W.C., Deshazer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T., Ulrich R.L., Ronning C.M., Brinkac L.M., Daugherty S.C.,
RA Davidson T.D., Deboy R.T., Dmitrov G., Dodson R.J., Durkin A.S.,
RA Gwinn M.L., Haft D.H., Khouri H., Kolonay J.F., Madupu R.,
RA Mohammed Y., Nelson W.C., Radune D., Romero C.M., Sarria S.,
RA Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y., Zafar N.,
RA Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251(2004).
```

```
DR EMBL; CP000011; AAU45430.1; -.
SQ SEQUENCE 2338 AA; 245390 MW; 5E0725F06ED4740 CRC64;

Query Match 23.4%; Score 107; DB 2; Length 2338;
Best Local Similarity 53.7%; Pred. No. 1.3;
Matches 29; Conservative 3; Mismatches 20; Indels 2; Gaps 1;

QY 18 SAQNPTTAADT--YPATGPADDEAPDAETTAATAATTATTAATTAATTAATTA 69
Db 1290 SSSSATAATAATAATTATTAATTATTTATTTATTTATTTATTTATTTATTTA 1343

RESULT 14
Q6ML46
ID Q6ML46 PRELIMINARY; PRT; 71 AA.
AC Q6ML46;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein precursor.
GN OrderedLocusNames=Bd2175;
OS Bdellovibrio bacteriovorus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;
OC Bdellovibrionaceae; Bdellovibrio.
OX NCBI_TaxID=959;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
RX PubMed=14752164; DOI=10.1126/science.1093027;
RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,
RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,
RA Sockett R.E., Schuster S.C.;
RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
genomic perspective.";
RL Science 303:689-692(2004).
DR EMBL; BX842651; CAE80011.1; -.
KW Complete proteome.
FT SIGNAL 1 22 Potential.
SQ SEQUENCE 71 AA; 6827 MW; 043CA16997EBC129 CRC64;

Query Match 23.3%; Score 106.5; DB 2; Length 71;
Best Local Similarity 46.5%; Pred. No. 0.054;
Matches 33; Conservative 3; Mismatches 30; Indels 5; Gaps 3;

QY 1 MKFLAVLVLLGVSI---FLVSAQNPTTAADTPATGPDADDEAPDAETTAAT-TATTA 56
Db 1 MKAIIVVAALLALSFTTGFTCSKNQPET-APATAPAAEATATAPAEPTAATETAPAA 59

QY 57 APTTATTAAT 67
Db 60 APTAAPAGET 70

RESULT 15
Q91X55
ID Q91X55 PRELIMINARY; PRT; 208 AA.
AC Q91X55;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Salivary protein 2.
GN Name=Spt2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Salivary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueno T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Salivary gland;  
 RA Strauberg R.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC012233; AAH12233.1; -;  
 DR MGI; MGI:98393; Cpt2.  
 DR GO; GO:0005615; C:extracellular space; TAS  
 SQ SEQUENCE 208 AA; 22895 MW; FCR2424A2898CA79 CRC64;  
 Query Match 23.3%; Score 106.5; DB 2; Length 208;  
 Best Local Similarity 43.2%; Pred. No. 0.15;  
 Matches 32; Conservative 9; Mismatches 24; Indels 9; Gaps 3;  
 Qy 1 MKFLAVLVLCVSYFLYSQAQPTTAAPADYTPANGPADDEAPDAET--TAAATTTATTAAP 58  
 Db 1 MKFRALLVLLGVSYFLYSQAQPTTAAPADYTPANGPADDEAPDAET--TAAATTTATTAAP 56  
 Qy 59 TTATTAATTTARKD 72  
 Db 57 TQAD---STVENQD 67  
 RESULT 16  
 ID\_HUMAN STANDARD; PRT; 907 AA.  
 AC O60641; QNTV7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Clathrin coat assembly protein APL80 (Clathrin coat associated protein  
 DE APL80) (91 kDa synaptosomal-associated protein).  
 GN Name=SNAP91; Synonyms=KIAA0656;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Korani H., Nomura N., Ohara O.;  
 RA "Prediction of the coding sequences of unidentified human genes. X.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:169-176(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22935763; PubMed=14574404; DOI=10.1038/nature02055;  
 RA Mungall A.J., Palmer S.A., Sims S.K., Edwards C.A., Ashurst J.L.,  
 RA Wilming L., Jones M.C., Horton R., Hunt S.E., Scott C.E.,  
 RA Gilbert J.G.R., Clamp M.E., Bethel G., Milne S., Ainscough R.,  
 RA Almeida J.P., Ambrose K.D., Andrews T.D., Ashwell R.I.S.,  
 RA Babbage A.K., Bagguley C.L., Bailey J., Banerjee R., Barker D.J.,  
 RA Barlow K.F., Beare D.M., Beasley H., Beasley K., Bird C.P.,

RA Blakey S.E., Bray-Allen S., Brook J., Brown A.J., Brown J.Y.,  
 RA Burford D.C., Burrill W., Burton J., Carder C., Carter N.P.,  
 RA Chapman J.C., Clark S.Y., Clark C., Clegg S., Cobley V.,  
 RA Collier R.E., Collins J.E., Colman L.K., Corby N.R., Coville G.J.,  
 RA Culley K.M., Dhali P., Davies J., Dunn M., Earthrowl M.E.,  
 RA Ellington A.E., Evans K.A., Faulkner L., Francis M.D., Frankish A.,  
 RA Frankland J., French L., Garner P., Garnett J., Ghorri M.J.,  
 RA Gilby L.M., Gillson C.J., Glithero R.J., Grafham D.V., Grant M.,  
 RA Gribble S., Griffiths C., Griffiths M.N.D., Hall R., Halls K.S.,  
 RA Hammond S., Harley J.L., Hart E.A., Heath P.D., Heathcott R.,  
 RA Holmes S.J., Howden P.J., Howe K.L., Howell G.R., Huckle E.,  
 RA Humphray S.J., Humphries M.D., Hunt A.P., Johnson C.M., Joy A.A.,  
 RA Kay M., Keenan S.J., Kimberley A.M., King A., Laird G.K., Langford C.,  
 RA Lawlor S., Leongamornlert D.A., Leversha M., Lloyd C.R., Lloyd D.M.,  
 RA Loveland J.E., Lovell J., Martin S., Mashreghi-Mohammadi M.,  
 RA McLaren G.I., Matthews L., McCann O.T., McLaren S.J., McLeay K.,  
 RA McMurray A., Moore M.J.F., Mullikin J.C., Niblett D., Nickerson T.,  
 RA Novik K.L., Oliver K., Overton-Larty E.K., Parker A., Patel R.,  
 RA Pearce A.V., Peck A.I., Phillimore B.J.C.T., Phillips S., Plumb R.W.,  
 RA Porter K.M., Ramsey Y., Ranby S.A., Rice C.M., Ross M.T., Searle S.M.,  
 RA Sehra H.K., Sheridan E., Skuce C.D., Smith S., Smith M., Spraggon L.,  
 RA Squares S.L., Steward C.A., Sycamore N., Tamlyn-Hall G., Tester J.,  
 RA Theaker A.J., Thomas D.W., Thorpe A., Tracey A., Tromans A., Tubby B.,  
 RA Wall M., Wallis J.M., West A.P., White S.S., Whitehead S.L.,  
 RA Whittaker H., Wild A., Willey D.J., Wilmer T.E., Wood J.M., Wray P.W.,  
 RA Wyatt J.C., Young L., Younger R.M., Bentley D.R., Coulson J., Beck S.;  
 RA Durbin R., Hubbard T., Sulston J.E., Dunham I., Rogers J., Beck S.;  
 RT "The DNA sequence and analysis of human chromosome 6.";  
 RL Nature 425:805-811(2003).  
 RN [3]  
 RP SEQUENCE OF 794-907 FROM N.A.  
 RC TISSUE=Brain;  
 RA Yu W., Gibbs R.A.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Adaptins are components of the adaptor complexes which  
 CC link clathrin to receptors in coated vesicles. Clathrin-associated  
 CC protein complexes are believed to interact with the cytoplasmic  
 CC tails of membrane proteins, leading to their selection and  
 CC concentration. Binding of APL80 to clathrin triskelia induces  
 CC their assembly into 60-70 nm coats (By similarity).  
 CC -!- SUBUNIT: Binds AP2A2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the  
 CC cytoplasmic face of coated vesicles in the plasma membrane (By  
 CC similarity).  
 CC -!- DOMAIN: Possesses a three domain structure: the N-terminal 300  
 CC residues harbor a clathrin binding site, an acidic middle domain  
 CC 450 residues, interrupted by an Ala-rich segment, and the C-  
 CC terminal domain (166 residues).  
 CC -!- PTM: Phosphorylated (By similarity).  
 CC -!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).  
 -----  
 DR EMBL; AB014556; BAA31631.2; ALT\_INIT.  
 DR EMBL; AL109915; CAB89292.1; -;  
 DR EMBL; AF054993; AAC09352.1; -;  
 DR HSP; O55011; IHG5.  
 DR Genew; HGNC:14986; SNAP91.  
 DR MIM; 607923; -;  
 DR InterPro; IPR008943; PI\_bind\_N.  
 DR Pfam; PF01417; ENTH; 1.  
 DR PROSITE; PS00942; ENTH; 1.  
 KW Coated pits; phosphorylation.  
 FT DOMAIN 14 145 ENTH.  
 FT DOMAIN 361 583 Ala-rich.  
 FT DOMAIN 536 556 Thr-rich.  
 FT DOMAIN 809 897 Pro-rich.

```

FT CONFLICT 293 295 NEG -> KYA (in Ref. 2).
FT CONFLICT 382 386 Missing (in Ref. 2).
SQ SEQUENCE 907 AA; 92501 MW; 23959C2B54F5EBF1 CRC64;

Query Match 23.0%; Score 105.5; DB 1; Length 907;
Best Local Similarity 48.3%; Pred. No. 0.72;
Matches 29; Conservative 5; Mismatches 23; Indels 3; Gaps 2;

Qy 24 TAAPADTYPATGPADDEAPDAETTAATAATTATAAPTATTATAASTTARKDIPVLKRWGDL 83
Db 511 TASTAPPVPATAPS--PAPAVAAAAATTAATAATTTTTTSAATATAPPALDIP-GDL 567

RESULT 17
Q84L71 PRELIMINARY; PRT; 184 AA.
AC 084L71;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Elicitin protein.
GN Name=SOJ6;
OS Phytophthora sojae.
OC Eukaryota; Stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC Phytophthora.
ON NCBI_TaxID=67593;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P6497; TISSUE=Mycelia;
RA Qutob D., Hultena E., Gijzen M., Kamoun S.;
RT "Variation in structure and activity among elicitors from Phytophthora sojae.";
RL Mol. Plant Pathol. 4:119-124(2003).
DR EMEL; AY183412; AA024643.1; -.
DR HSP; P15569; ILJP.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0006952; P:defense response; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR02200; Elicitin.
DR InterPro; IPR006970; PT.
DR Pfam; PF00964; Elicitin; 1.
DR Pfam; PF04886; PT; 2.
DR PRINTS; PR00948; ELICITIN.
DR ProDom; PD004802; Elicitin; 1.
SQ SEQUENCE 184 AA; 18430 MW; 22814039022E5D4C CRC64;

Query Match 22.9%; Score 105; DB 2; Length 184;
Best Local Similarity 46.0%; Pred. No. 0.18;
Matches 29; Conservative 5; Mismatches 21; Indels 8; Gaps 3;

Qy 21 NPTTAAADTYPATGPAD-----DEAPDAETTAATAATTATAAPTATTATAASTTARKDIP 74
Db 120 NP-TSAPDA-PTSAPTADPTDPTSDPTDPTSDPTDPTSDPTDPTSDPTDPTSDPTDPTSDPT 177

Qy 75 VLP 77
Db 178 VVP 180

RESULT 18
Q73T86 PRELIMINARY; PRT; 565 AA.
AC Q73T86
DT 05-JUN-2004 (TrEMBLrel. 27, Created)
DT 05-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OtherLocusNames=MAP3832c;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1770;
RX [1]

```

Matches 26; Conservative 11; Mismatches 32; Indels 10; Gaps 1

QY 4 L A V L V L G V S I F L V S A Q N P T T A A P A D Y P A T G P A D D E A ----- P D A E T T A A A T T A 53  
 Db 6 L L I A L A M G A T A A C A C G T K T P E T A Q D S N P A S N P A A N E A Q A A D Q A A A N P P A D A T P A A A D T 65

QY 54 T T A A P T A T T A A S T A R K D 72  
 Db 66 A A A A N T A A D A A A T T T P P A D 84

RESULT 20  
 Q81Q88 PRELIMINARY; PRT; 130 AA.

AC Q81Q88;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CG32077-PA.  
 GN Name=nol; ORFNames=CG32077;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]

SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abiril J.F., Agbayani A., An H.J., Andrews-Pfaunkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,  
 RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.S., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,  
 RA Rainert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodager, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [2]

SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,

[illegible]

```
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RT repetitive region.";
RL Biochem. J. 358:763-772(2001).
DR EMBL; AJ298317; CAC83674.1; -.
DR HSSP; P56682; 1CCV.
DR Genew; HGNC:7515; MUC5AC.
DR Pfam; PF01826; TIL; 2.
DR Pfam; PF00094; VWD; 3.
DR SMART; SM00214; VWC; 3.
DR SMART; SM00216; VWD; 3.
DR PROSITE; PS00527; RIBOSOMAL_S14; UNKNOWN 2.
FT NON TER 2448
SQ SEQUENCE 2448 AA; 260968 MW; A5C1BD627844D952 CRC64;

Query Match 22.6%; Score 103.5; DB 2; Length 2448;
Best Local Similarity 43.4%; Pred. No. 2.7;
Matches 23; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

Qy 23 TTAP-ADTYPATGPADDEAPDAETTAATTAATTAAPTTATTAASTTARKDIP 74
Db 2396 TTSAPSTSGPGTSPVFTTSITSAPTTSITSAPTTSITSAPTTSITSAPTTSAP 2448

RESULT 24
Q68DL9 PRELIMINARY; PRT; 816 AA..
AC Q68DL9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp78100519.
GN Name=DKFZp78100519;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RG The German cDNA Consortium;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the ENBL/GenBank/DBSJ databases.
DR EMBL; CR749348; CAH18201.1; -.
DR InterPro; IPR011417; ANTH.
DR InterPro; IPR001026; Epsin N.
DR InterPro; IPR008943; PI_bind_N.
DR Pfam; PF07651; ANTH; 1.
DR Pfam; PF01417; ENTH; 1.
DR SMART; SM00273; ENTH; 1.
KW Hypothetical protein.
SQ SEQUENCE 816 AA; 83398 MW; 3E39EBC4F1EA965E CRC64;

Query Match 22.5%; Score 103; DB 2; Length 816;
Best Local Similarity 49.1%; Pred. No. 1.1;
Matches 26; Conservative 4; Mismatches 21; Indels 2; Gaps 1;

Qy 24 TAAPADTYPATGPADDEAPDAETTAATTAATTAAPTTATTAASTTARKDIPVL 76
Db 474 TASTAPPVPATAPS--PAPAVAAATAATAATAATTTTTSATATTAPPAL 524

RESULT 25
Q25645 PRELIMINARY; PRT; 1787 AA.
AC Q25645;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein-1.
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX NCBI_TaxID=5821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K173;
RX MEDLINE=97235961; PubMed=9080880;
RA Toebe C.S., Clements J.D., Cardenas L., Jennings G.J., Wiser M.F.;
RT "Evaluation of immunogenicity of an oral Salmonella vaccine expressing
RT recombinant Plasmodium berghei merozoite surface protein-1.";
RL Am. J. Trop. Med. Hyg. 56:192-199(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K173;
RX MEDLINE=98324404; PubMed=9662027; DOI=10.1016/S0166-6851(98)00016-4;
RA Jennings G.J., Toebe C.S., van Belkum A., Wiser M.F.;
RT "The complete sequence of Plasmodium berghei merozoite surface
RT protein-1 and its inter- and intra-species variability.";
RL Mol. Biochem. Parasitol. 93:43-55(1998).
DR EMBL; U43521; AAC28871.1; -.
DR HSSP; Q9GSQ9; IN11.
DR InterPro; IPR010901; MSP1_C.
DR Pfam; PF07462; MSP1_C; 1.
KW Merozoite.
SQ SEQUENCE 1787 AA; 198155 MW; A585D64F5148E4EA CRC64;

Query Match 22.4%; Score 102.5; DB 2; Length 1787;
Best Local Similarity 49.2%; Pred. No. 2.4;
Matches 32; Conservative 5; Mismatches 19; Indels 9; Gaps 4;

Qy 18 SAQN--PTTAAPADTYPATGP--ADDEAPDAETTAATTAATTAASTTARKDI 73
Db 778 SSQNPPTTAAT---PATTPEAATAATPATTPEAATTTTSTTTSTT-TSTTTSTTT 832
Qy 74 PVLPK 78
Db 833 PVMTK 837

Search completed: September 26, 2005, 08:30:12
Job time : 96.967 secs
```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:44:37 ; Search time 96.0659 Seconds  
(without alignments)  
495.736 Million cell updates/sec

Title: US-09-975-502A-5

Perfect score: 475

Sequence: 1 MKLLMVLMLAALSQHCYAGS.....LSNVFPMQLIYDSSLCDLF 93

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Uniprot 03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	%	Query	Length	DB	ID	Description
1	475	100.0	93	1	MGBA	HUMAN	Q13296	homo sapien
2	417	87.8	120	2	Q6NX70		Q6NX70	homo sapien
3	281	59.2	95	1	MGBB	HUMAN	Q75556	homo sapien
4	272	57.3	93	2	Q6UWM4		Q6UWM4	homo sapien
5	272	57.3	93	2	Q9GK63		Q9GK63	oryctolagus
6	252	53.1	93	2	Q9GK61		Q9GK61	oryctolagus
7	249	52.4	93	2	Q863D2		Q863D2	sus scrofa
8	240	50.5	93	2	Q9GK64		Q9GK64	oryctolagus
9	227	47.8	93	2	Q9GK62		Q9GK62	oryctolagus
10	204.5	43.1	95	1	PSC3	RAT	P02780	rattus norv
11	196.5	41.4	95	2	Q9JHB9		Q9JHB9	rattus norv
12	163.5	34.4	94	2	Q9QXF3		Q9QXF3	mesocricetu
13	156.5	32.9	94	2	Q9QXF2		Q9QXF2	mesocricetu
14	86.5	18.2	1232	2	Q6BEP7		Q6BEP7	homo sapien
15	79.5	16.7	1282	2	Q7LG62		Q7LG62	homo sapien
16	79.5	16.7	1302	2	Q9H582		Q9H582	homo sapien
17	79.5	16.7	1319	2	Q6P446		Q6P446	homo sapien
18	78.5	16.5	501	2	Q7P617		Q7P617	fusobacteri
19	78.5	16.5	529	2	Q6F1Z9		Q6F1Z9	mesoplasma
20	78	16.4	91	1	UTER	LEPCA	P06913	lepus capen
21	78	16.4	512	2	Q9ZW95		Q9ZW95	arabidopsis
22	77	16.2	634	2	Q6CK74		Q6CK74	kluveromyc
23	76.5	16.1	343	1	VA0D	SCHPO	O13753	schizosacch
24	76.5	16.1	2169	2	Q7RHE7		Q7RHE7	plasmodium
25	76	16.0	91	1	UTER	RABIT	P02779	oryctolagus
26	75.5	15.9	807	2	Q8IIIR1		Q8IIIR1	plasmodium
27	74.5	15.7	345	2	Q6M0V5		Q6M0V5	methanococc
28	74.5	15.7	842	2	Q6FR44		Q6FR44	candida gla
29	74	15.6	521	2	Q47510		Q47510	escherichia
30	74	15.6	1149	1	HDA6	MOUSE	Q92245	mus musculu
31	74	15.6	1464	2	Q8IHS3		Q8IHS3	plasmodium

ALIGNMENTS

32	74	15.6	2870	2	Q76M60		Q76M60	melon yello
33	73.5	15.5	309	2	P93550		P93550	spinacia ol
34	73.5	15.5	493	2	Q832K8		Q832K8	enterococc
35	73	15.4	261	2	Q8QM27		Q8QM27	cowpox viru
36	73	15.4	421	2	Q93MD1		Q93MD1	clostridium
37	73	15.4	629	2	Q9CAM8		Q9CAM8	arabidopsis
38	73	15.4	1150	1	SCC3	YEAST	SCC3	saccharomyc
39	72.5	15.3	93	2	Q65C83		Q65C83	neotomon
40	72.5	15.3	266	2	Q8REY4		Q8REY4	fusobacteri
41	72.5	15.3	270	2	Q9LZW2		Q9LZW2	arabidopsis
42	72.5	15.3	282	2	Q9ZFP8		Q9ZFP8	haemophilus
43	72.5	15.3	446	2	Q7PBG1		Q7PBG1	rickettsia
44	72.5	15.3	481	2	Q6VBJ5		Q6VBJ5	capsicum an
45	72.5	15.3	651	2	Q9UY59		Q9UY59	pyrococcus
46	72.5	15.3	728	2	Q8RG12		Q8RG12	fusobacteri
47	72	15.2	91	1	UTER	HUMAN	P11684	homo sapien
48	72	15.2	271	2	Q8TTV0		Q8TTV0	methanosarc
49	72	15.2	590	2	Q66BK9		Q66BK9	yersinia ps
50	72	15.2	849	2	Q8ISV7		Q8ISV7	plasmodium
51	72	15.2	985	2	Q8I159		Q8I159	plasmodium
52	72	15.2	1440	2	Q96279		Q96279	plasmodium
53	72	15.2	1527	2	Q8IM29		Q8IM29	plasmodium
54	71.5	15.1	136	2	Q7R6A8		Q7R6A8	giardia lam
55	71.5	15.1	287	2	Q8DI32		Q8DI32	synechococc
56	71.5	15.1	313	2	Q84EX1		Q84EX1	lactobacill
57	71.5	15.1	333	2	Q9JFV8		Q9JFV8	canine herp
58	71.5	15.1	728	2	Q7PAN7		Q7PAN7	fusobacteri
59	71.5	15.1	759	2	Q9ZK71		Q9ZK71	helicobacte
60	71	14.9	295	2	Q7VF62		Q7VF62	helicobacte
61	71	14.9	395	2	Q93N60		Q93N60	coxiella bu
62	71	14.9	413	2	Q757A7		Q757A7	ashbya goss
63	71	14.9	420	2	Q83DM0		Q83DM0	coxiella bu
64	71	14.9	1152	2	Q8CGC3		Q8CGC3	mus musculu
65	71	14.9	2788	2	Q7RI61		Q7RI61	plasmodium
66	70.5	14.8	191	2	Q6MUE2		Q6MUE2	mycoplasma
67	70.5	14.8	331	1	PE60	ARATH	PE60	arabidopsis
68	70.5	14.8	626	2	Q6LB53		Q6LB53	oligotropha
69	70.5	14.8	641	2	Q86AQ2		Q86AQ2	dictyosteli
70	70.5	14.8	1628	2	Q7RA78		Q7RA78	plasmodium
71	70	14.7	159	2	Q8JN13		Q8JN13	infectious
72	70	14.7	171	2	Q6KGE7		Q6KGE7	bacterioph
73	70	14.7	493	2	Q6AT25		Q6AT25	oryza sativ
74	70	14.7	561	1	EST5	RAT	EST5	rattus norv
75	70	14.7	561	2	Q6AYX3		Q6AYX3	rattus norv
76	70	14.7	697	2	Q9U921		Q9U921	tetrahymena
77	70	14.7	738	2	Q8EUN0		Q8EUN0	mycoplasma
78	70	14.7	780	2	Q6CNK7		Q6CNK7	kluveromyc
79	69.5	14.6	174	2	Q7PE57		Q7PE57	anopheles g
80	69.5	14.6	187	1	DEF	FREDI	P94601	fremyella d
81	69.5	14.6	270	2	Q8YXC0		Q8YXC0	anabaena sp
82	69.5	14.6	320	2	Q8IET2		Q8IET2	plasmodium
83	69.5	14.6	344	2	Q8CR97		Q8CR97	staphylococ
84	69.5	14.6	406	2	Q97IL0		Q97IL0	clostridium
85	69.5	14.6	685	2	Q75Y87		Q75Y87	streptococc
86	69.5	14.6	2766	2	Q9QZ88		Q9QZ88	rattus norv
87	69	14.5	90	2	Q9GK67		Q9GK67	oryctolagus
88	69	14.5	144	1	IL3	BOVIN	P49875	bos taurus
89	69	14.5	160	2	Q8QV00		Q8QV00	infectious
90	69	14.5	170	2	Q66QJ7		Q66QJ7	borrelia ga
91	69	14.5	205	2	Q05154		Q05154	hepatitis c
92	69	14.5	441	1	COAT	SOCMV	P15627	soybean chl
93	69	14.5	475	2	Q81588		Q81588	hepatitis c
94	69	14.5	510	2	Q87UG6		Q87UG6	pseudomonas
95	69	14.5	625	2	Q01976		Q01976	caenorhabdi
96	69	14.5	643	2	Q9NWL6		Q9NWL6	homo sapien
97	69	14.5	644	2	Q80936		Q80936	human papil
98	69	14.5	647	2	Q8EVR1		Q8EVR1	mycoplasma
99	69	14.5	1021	2	Q25200		Q25200	helicobacte
100	69	14.5	3033	2	Q7T7J0		Q7T7J0	hepatitis c

```
RESULT 1
MGBA_HUMAN
ID MGBA_HUMAN STANDARD; PRT; 93 AA.
AC Q13256; Q86WH8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 05-JUL-2004 (Rel. 44, last annotation update)
DE Mammaglobin A precursor (Mammaglobin 1) (Secretoglobin family 2A member 2).
GN Name=SCGB2A2; Synonym=MGb1, UGB2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast;
RX MEDLINE=96223698; PubMed=8631025;
RA Watson M.A., Fleming T.P.;
RT "Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer.";
RL Cancer Res. 56:860-865(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=98147371; PubMed=9488047; DOI=10.1038/sj.onc.1201597;
RA Watson M.A., Darrow C., Zimonjic D.B., Popescu N.C., Fleming T.;
RT "Structure and transcriptional regulation of the human mamaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13.";
RL Oncogene 16:817-824(1998).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Zhao L., Nan K.;
RT "An alternative splicing isoform of mammaglobin.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q13296-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q13296-2; Sequence=VSP_009122;
CC -!- TISSUE SPECIFICITY: Mammary gland specific. Over-expressed in breast cancer.
CC -!- SIMILARITY: Belongs to the uteroglobin family. Lipophilin subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U33147; AAC50375.1; -.
CC EMBL; AF015224; AAC39608.1; -.
CC EMBL; AV217100; AAO6011.1; -.
CC Genew; HGNC:7050; SCGB2A2.
CC MIM; 605562; -.
CC InterPro; IPR003627; Mambg/prostatn.
CC InterPro; IPR006039; Utergl.
CC InterPro; IPR000329; Uteroglobin subf.
CC InterPro; IPR006038; Uteroglobin_supf.
CC Pfam; PF01099; Uteroglobin; 1.
CC ProDom; PD029354; Mambg/prostatn; 1.
CC SMART; SM00096; UTG; 1.
CC PROSITE; PS00403; UTEROGLIBIN 1; FALSE_NEG.
CC PROSITE; PS00404; UTEROGLIBIN 2; 1.
CC Alternative splicing; Glycoprotein; Signal.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 93 Mammaglobin A.
FT CARBOHYD 53 53 N-linked (GlcNAc..) (Potential).
FT
```

```
FT CARBOHYD 68 68 N-linked (GlcNAc..) (Potential).
FT VARSPLIC 79 81 Missing (in isoform 2).
FT /FTID=VSP_009122.
SQ SEQUENCE 93 AA; 10499 MW; 2896E8C43BF053E2 CRC64;
Query Match 100.0%; Score 475; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.5e-37;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLLMVLMLAALSOHCYAGSGCPLENNVISTKINPOVSKTEYKELQEFIDNATTNAID 60
Db 1 MKLLMVLMLAALSOHCYAGSGCPLENNVISTKINPOVSKTEYKELQEFIDNATTNAID 60
QY 61 ELKECFLNQTDETLSNVFVFMQLIYDSSLCDLF 93
Db 61 ELKECFLNQTDETLSNVFVFMQLIYDSSLCDLF 93
RESULT 2
Q6NX70
ID Q6NX70 PRELIMINARY; PRT; 120 AA.
AC Q6NX70;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin and meninges pool- skin;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smaluk D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin and meninges pool- skin;
RA Strausberg R.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC067220; AAH67220.1; -.
DR GO; GO:0005496; F:steroid binding; IEA.
DR InterPro; IPR003627; Mambg/prostatn.
DR InterPro; IPR006038; Uteroglobin_supf.
DR Pfam; PF01099; Uteroglobin; 1.
DR ProDom; PD029354; Mambg/prostatn; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 120 AA; 13289 MW; 4A37A6296CE9039B CRC64;
Query Match 87.8%; Score 417; DB 2; Length 120;
Best Local Similarity 94.3%; Pred. No. 6.6e-32;
Matches 82; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 MKLLMVLMLAALSOHCYAGSGCPLENNVISTKINPOVSKTEYKELQEFIDNATTNAID 60
```

Db 1 M K L L M V L M L A A L S O H C V A G S G C P L L E N V I S K T I N P Q V S K Y E K L L Q E F I D D N A T T W A I D 60  
Qy 61 E L K E C F L N Q T D E T L S N V E F M Q L I Y D S 87  
Db 61 E L K E C F L N Q T D E T L S N V E F M Q I S F S S 87  
RESULT 3  
MGBB HUMAN STANDARD; PRT; 95 AA.  
AC 075556;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Mamaglobin B precursor (Mamaglobin 2) (Lipophilin C) (lacryglobin)  
DE (Secretoglobin family 2A member 1).  
GN Name=SCGB2A1; Synonyms=LIPH, MGB2, UGB3;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99026127; PubMed=9806831; DOI=10.1006/geno.1998.5539;  
RA Becker R.W., Darrow C., Zimonjic D.B., Popescu N.C., Watson M.A.,  
Fleming T.P.;  
RT "Identification of mamaglobin B, a novel member of the uterogloblin  
gene family";  
RL Genomics 54:70-78 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99167354; PubMed=10066439; DOI=10.1006/bbrc.1999.0274;  
RA Zhao C., Nguyen T., Yusifov T., Glasgow B.J., Lehrer R.I.;  
RT "Lipophilins: human peptides homologous to rat prostatein";  
RL Biochem. Biophys. Res. Commun. 256:147-155 (1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thyroid;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Haile F.,  
Daplatenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
Bohak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Vallalao D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Ridguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [4]  
RP SEQUENCE OF 19-85.  
RC TISSUE=Tears;  
RX MEDLINE=98163342; PubMed=9504814;  
RA Molloy M.P., Bolis S., Herbert B.R., Ou K., Tyler M.I., van Dyk D.D.,  
Willcox M.D., Gooley A.A., Williams K.L., Morris C.A., Walsh B.J.;  
RT "Establishment of the human reflex tear two-dimensional polyacrylamide  
gel electrophoresis reference map: new proteins of potential  
diagnostic value";  
RL Electrophoresis 18:2811-2815 (1997).  
RN [5]  
RP SEQUENCE OF 19-46 AND 60-78, AND MASS SPECTROMETRY.  
RC TISSUE=Tears;

RX MEDLINE=98385871; PubMed=9720917; DOI=10.1016/S0014-5793(98)00852-7;  
RA Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Qu X.-D., Martin D.,  
Glasgow B.J.;  
RT "Lipophilin, a novel heterodimeric protein of human tears.";  
RL FEBS Lett. 432:163-167 (1998).  
CC -I- FUNCTION: May bind androgens and other steroids, may also bind  
estradiol, a chemotherapeutic agent used for prostate cancer.  
CC May be under transcriptional regulation of steroid hormones.  
CC -I- SUBUNIT: Heterodimer of a lipophilin A and a lipophilin C  
(mamaglobin B) monomer associated head to head.  
CC -I- TISSUE SPECIFICITY: Expressed in thymus, trachea, kidney, steroid  
responsive tissues (prostate, testis, uterus, breast and ovary)  
CC and salivary gland.  
CC -I- MASS SPECTROMETRY: MW=8854.94; METHOD=Electrospray; RANGE=19-95;  
CC NOTE=Ref.5.  
CC -I- SIMILARITY: Belongs to the uterogloblin family. Lipophilin  
subfamily.  
CC -----  
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
CC EMBL; AF071219; AAC79996.1; -  
CC EMBL; AJ224173; CAA11865.1; -  
CC EMBL; BC062218; AAH62218.1; -  
CC Genbank; HGNC:7051; SCGB2A1.  
CC MIM; 604398; -  
CC GO; GO:0005497; P: androgen binding; NAS.  
CC InterPro; IPR003627; Mamgb/prostatn.  
CC InterPro; IPR000329; Uterogloblin subf.  
CC InterPro; IPR006038; Uterogloblin subf.  
CC Pfam; PF01099; Uterogloblin; 1.  
CC PRODOM; PD029354; Mamgb/prostatn; 1.  
CC PROSITE; PS00403; UTEROGLOBIN\_1; FALSE NEG.  
CC PROSITE; PS00404; UTEROGLOBIN\_2; FALSE NEG.  
KW Direct protein sequencing; Glycoprotein; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 95 Mamaglobin B.  
FT CARBOHYD 68 68 N-linked (GlcNAc...) (Potential).  
SQ SEQUENCE 95 AA; 10884 MW; 0719738289F89F8D CRC64;  
Query Match 59.2%; Score 281; DB 1; Length 95;  
Best Local Similarity 58.7%; Pred. No. 4.4e-19;  
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;  
Qy 1 M K L L M V L M L A A L S O H C V A G S G C P L L E N V I S K T I N P Q V S K Y E K L L Q E F I D D N A T T W A I D 60  
Db 1 M K L L M V L M L A A L L H C V A D S G C K L L E D M V K T I N S D I S I P E Y K E L L Q E F I D S D A A A E A M G 60  
Qy 61 E L K E C F L N Q T D E T L S N V E F M Q L I Y D S S L C D L 92  
Db 61 K F Q C F L N Q S H R T L K N F L M M H T V I D S I W C N M 92  
RESULT 4  
Q6UWM4 PRELIMINARY; PRT; 93 AA.  
ID Q6UWM4  
AC Q6UWM4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE KVM3106.  
GN ORFNames=UQ03106;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.



```
Matches 52; Conservative 10; Mismatches 31; Indels 0; Gaps 0;
QY 1 M K L L V M L M L A A L S O H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60
D b 1 M K L A M L M L V A L P V Y C A G S G C S Y L E R V I S D T S S V T T D V I L A S L Q E Y I S S D T T Q A I K 60
QY 61 E L K E C F L N Q T D E T L S N V E F M Q L I Y D S S L C D L F 93
D b 61 E L R E C F L K Q S E E T L E N F S V F M Q V T Y N S K L C A A F 93

RESULT 8
Q9GK64 PRELIMINARY; PRT; 93 AA.
AC Q9GK64
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lipophilin CL.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lacrimal gland;
RA Zhao C., Nguyen T.X., Lehrer R.I.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF308617; AAG42805.1; -.
DR GO; GO:0005496; F:steroid binding; IEA.
DR InterPro; IPR003627; Mamgb/prostatn.
DR InterPro; IPR006038; Uteroglobulin_supf.
DR Pfam; PF01099; Uterogloblin; 1.
DR ProDom; PD029354; Mamgb/prostatn; 1.
SQ SEQUENCE 93 AA; 10350 MW; 9F9206C44372804D CRC64;

Query Match 50.58; Score 240; DB 2; Length 93;
Best Local Similarity 50.08; Pred. No. 3.4e-15;
Matches 45; Conservative 23; Mismatches 22; Indels 0; Gaps 0;
QY 1 M K L L V M L M L A A L S O H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60
D b 1 M K W I L V L A T L P F Y C Y A G S G C V I L E S V L D K T I D P S V S V E D Y T Y L Q K Y I L T D A A K V A L E 60
QY 61 E L K E C F L N Q T D E T L S N V E F M Q L I Y D S S L C 90
D b 61 E L K Q C F L S Q S N E T L A N K V L E Y A V F D S L Y C 90

RESULT 9
Q9GK62 PRELIMINARY; PRT; 93 AA.
AC Q9GK62
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lipophilin CP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Zhao C., Nguyen T.X., Lehrer R.I.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF308619; AAG42807.1; -.
DR GO; GO:0005496; F:steroid binding; IEA.
DR InterPro; IPR003627; Mamgb/prostatn.
DR InterPro; IPR006038; Uterogloblin_supf.
DR Pfam; PF01099; Uterogloblin; 1.
DR ProDom; PD029354; Mamgb/prostatn; 1.
SQ SEQUENCE 93 AA; 10332 MW; C9DC35B17D372F32 CRC64;
```

```
Query Match 47.88; Score 227; DB 2; Length 93;
Best Local Similarity 43.3%; Pred. No. 5.8e-14;
Matches 39; Conservative 25; Mismatches 26; Indels 0; Gaps 0;
QY 1 M K L L V M L M L A A L S O H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60
D b 1 M K C V I A L M L A A L P L Y C Y A G S G C Q L L D M V T K I L D S Q I S L T D Y H N F F K N L S S G A A E M A V K 60
QY 61 E L K E C F L N Q T D E T L S N V E F M Q L I Y D S S L C 90
D b 61 D F K Q C F L M Q S N E T L N N I K V F L E T V Y N S P F C 90
```

```
RESULT 10
PSC3_RAT STANDARD; PRT; 95 AA.
ID PSC3_RAT
AC P02780; Q63463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Prostatic steroid-binding protein C3 chain precursor (Prostatein peptide C3).
DE Name=Psbpc3;
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83082848; PubMed=6294095;
RA Parker M.G., White R., Hurst H., Needham M., Tilly R.;
RT "Prostatic steroid-binding protein. Isolation and characterization of C3 genes.";
RT J. Biol. Chem. 258:112-15(1983).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83238526; PubMed=6190812;
RA Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M., French P.S.;
RT "Isolation of two genomic sequences encoding the Mr = 14,000 subunit of rat prostatein.";
RL J. Biol. Chem. 258:8861-8866(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92165796; PubMed=1537831;
RA Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M., French P.S.;
RT "Response elements of the androgen-regulated C3 gene.";
RL J. Biol. Chem. 267:4456-4466(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=92218467; PubMed=1339454;
RA Tan J.A., Marschke K.B., Ho K.C., Perry S.T., Wilson E.M., French P.S.;
RL J. Biol. Chem. 267:7958-7958(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=84057754; PubMed=6685625;
RA Hurst H.C., Parker M.G.;
RT "Rat prostatic steroid binding protein: DNA sequence and transcript maps of the two C3 genes.";
RL EMBO J. 2:769-774(1983).
RN [6]
RP SEQUENCE OF 19-95.
RX MEDLINE=81188769; PubMed=7014218;
RA Peeters B., Rombauts W., Mous J., Heyns W.;
RT "Structure of its glycosylated component C3.";
RL Eur. J. Biochem. 115:115-121(1981).
CC -I- FUNCTION: Part of prostatein which is the major secretory glycoprotein of ventral prostate gland. Steroid-binding protein; can bind non-polar steroids, cholesterol and a group of small
```



```
RESULT 13
ID Q9QXF2 PRELIMINARY; PRT; 94 AA.
AC Q9QXF2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Heteroglobin B2 subunit precursor.
GN Name=hgl.B2;
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Lake View; TISSUE=Harderian gland;
RC MEDLINE=21634881; PubMed=11684684; DOI=10.1074/jbc.M106678200;
RA Alvarez J., Vinas J., Alonso J.M.M., Albar J.P., Ashman K.,
RA Dominguez P.;
RT "Characterization and cloning of two isoforms of heteroglobin, a novel
RT heterodimeric glycoprotein of the secretoglobin-uteroglobin family
RT showing tissue-specific and sex differential expression.";
RL J. Biol. Chem. 277:233-242 (2002).
DR EMBL; AJ252139; CAB64661.1; -.
DR GO; GO:0005496; F:steroid binding; IEA.
DR InterPro; IPR003627; Mambg/prostatn.
DR InterPro; IPR006038; Uteroglobin_supf.
DR Pfam; PF01099; Uteroglobin; 1.
DR ProDom; PD029354; Mambg/prostatn; 1.
KW Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 94 AA; 10821 MW; 16C640C0674224C9 CRC64;

Query Match 32.9%; Score 156.5; DB 2; Length 94;
Best Local Similarity 34.1%; Pred. No. 2.9e-07;
Matches 31; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLLVLMALALSOHCYAGSGCPLELVISKINPOVSKTEYKELLQEFIDNATNNAI 59
Db 1 MKLVIVFLMAIIPYCVKTSNGGCRMDAIAKTINSVPMEYHETVKYKTTLPYIRSTV 60
QY 60 DELKECFLNQDTLSNVFVFMQLIYDSSLIC 90
Db 61 EKPECFAKQSNQTNIFVWYAVYNSDKC 91

RESULT 14
ID Q6BEP7 PRELIMINARY; PRT; 1232 AA.
AC Q6BEP7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Zinc finger motif Enhancer binding Protein-2.
GN Name=Zep-2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP Fukamizu A.;
RT "Zep-2: zinc finger motif enhancer binding protein-2.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB019255; BA032777.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 7.

RESULT 15
ID Q7LG67 PRELIMINARY; PRT; 1282 AA.
AC Q7LG67;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE KIAA1221 protein (Fragment).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345 (1999).
DR EMBL; AB033047; BAA86535.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
FT NON_TER
FT NON_TER
SQ SEQUENCE 1282 AA; 144490 MW; ED14FB14B1516A82 CRC64;

Query Match 16.7%; Score 79.5; DB 2; Length 1282;
Best Local Similarity 23.8%; Pred. No. 94;
Matches 21; Conservative 17; Mismatches 32; Indels 19; Gaps 3;

QY 6 VLMALALSOHCYAGSGCPLELVISKINPOVSKTEYKELLQEFIDNATNNAIDELK 65
Db 252 VSKITRYTEDCFSDSNC-----VPNKSQKQEV--DFLEQNEELQAVDSQKYA 296
QY 66 FL-----NQTDETLSNVFVFMQLIYDSSLIC 90
Db 297 LSKVKPESTDEDESVDFAFQHLIYNPDKC 325

RESULT 16
ID Q9H582 PRELIMINARY; PRT; 1302 AA.
AC Q9H582; Q9ULJ9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 26, Last annotation update)
DE DJ924G13.1 (KIAA1221 (Putative zinc finger protein)).
GN Name=dJ924G13.1;
```





```

OC Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;
OC Entomoplasmataceae; Mesoplasma.
OX NCBI_TaxID=2151;
RN [1]
RC STRAIN=Li / ATCC 33453;
RA Birren B.W., Stange-Thomann N., Hafez N., DeCaprio D., Fisher S.,
RA Butler J., Elkins T., Kodira C.D., Major J., Wang S., Nicol R.,
RA Nussbaum C.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AE017263; AAT75474.1; -.
DR GO; GO:0016021; C-integral to membrane; IEA.
DR GO; GO:0005524; F-ATP binding; IEA.
DR GO; GO:0042626; F-ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0000166; F-nucleotide binding; IEA.
DR GO; GO:0006810; P-transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011527; ABC_membrane_1.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00929; ABC_TMIF; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; UNKNOWN_1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 529 AA; 61918 MW; PF94FF44B473BE80B CRC64;

Query Match 16.5%; Score 78.5; DB 2; Length 529;
Best Local Similarity 25.8%; Pred. NO. 46;
Matches 24; Conservative 21; Mismatches 31; Indels 17; Gaps 4;

QY 4 LMTVLMALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLOEFFIDDDNATNAIDELK 63
DB 56 LRTTIIAIVSDYLSV-----LQNKIKIENQLRKKYDKILSDQFNSNIDTGG----- 105

QY 64 ECFLNQTDLSNVEVF-----MQLIYDSSLCDL 92
DB 106 --FINSSNKSISLEIFYYQISFLIF-SGICFL 135

RESULT 20
UTER LEPCA STANDARD; PRT; 91 AA.
AC P06913;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Uteroglobin precursor (Blastokinin).
GN Names=SCGB1A1; Synonyms=UGB, UGL;
OS Lepus capensis (Brown hare).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Lepus.
OX NCBI_TaxID=9981;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=86323069; PubMed=3019311;
RA Lopez de Haro M.S., Nieto A.;
RT "Nucleotide and derived amino acid sequences of a cDNA coding for pre-
RT uteroglobin from the lung of the hare (Lepus capensis).";
RL Biochem. J. 235:895-898(1986).
CC -1- FUNCTION: Uteroglobin binds progesterone specifically and with
CC high affinity. It may regulate progesterone concentrations
CC reaching the blastocyst. It is also a potent inhibitor of
CC phospholipase A2.
CC -1- SUBUNIT: Homodimer; antiparallel disulfide-linked.
CC -1- INDUCTION: Uteroglobin, synthesized in the uterus and lung, is
CC secreted by the uterus upon induction by progesterone.
CC -1- SIMILARITY: Belongs to the uteroglobin family.

```

---

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-aib.ch/announce/
CC or send an email to license@isb-aib.ch).
CC -----
DR EMBL; M25609; AAA30960.1; -.
DR PIR; A23825; UGRBL.
DR HSSP; P02779; IUTG.
DR InterPro; IPR006039; Utergl.
DR InterPro; IPR003628; Uteroglobin_sub.
DR InterPro; IPR000329; Uteroglobin_subf.
DR InterPro; IPR006038; Uteroglobin_subf.
DR Pfam; PF01099; Uteroglobin; 1.
DR PRINTS; PR00486; UTEROGLBIN.
DR ProDom; PD019935; Fel DI allergen; 1.
DR ProDom; PD012475; Uteroglobin_sub; 1.
DR SMART; SM00096; UTG; 1.
DR PROSITE; PS00403; UTEROGLBIN_1; 1.
DR PROSITE; PS00404; UTEROGLBIN_2; 1.
KW Phospholipase A2 inhibitor; Signal; Steroid-binding.
FT SIGNAL 1 21
FT CHAIN 22 91 Uteroglobin.
FT DISULFID 24 24 Interchain (with C-90).
FT DISULFID 90 90 Interchain (with C-24).
SQ SEQUENCE 91 AA; 9879 MW; 587614DAE9E4820F CRC64;

Query Match 16.4%; Score 78; DB 1; Length 91;
Best Local Similarity 28.1%; Pred. NO. 8.2;
Matches 27; Conservative 16; Mismatches 41; Indels 12; Gaps 3;

QY 1 MKLMLVLMALSOHCYAGSG--CP-----LLENVSKTINPQVSKTEYKELLOEFFIDDNA 54
DB 1 MKLTITLALVTLLALLSPASAGICPGFAHVNIENLLGT-----PSSYETSLKEFPDDA 54

QY 55 TTNVAIDELKECFNLQTDLSNVEFMQLIYDSSL 90
DB 55 MKDAGMQMKVLDLTPQTTRENIKLTKEIKVSPLC 90

RESULT 21
Q9ZW95 PRELIMINARY; PRT; 512 AA.
AC Q9ZW95;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE F5A8.3 protein (Atlg67110).
GN Names=F5A8.3;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,
RA Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Shinn P., Ecker J.R.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Ecker J.R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.

```

```
DR EMBL; AC004146; AAD10659.1; -.
DR EMBL; BT011622; AAS47628.1; -.
DR EMBL; BT012626; AAT06445.1; -.
DR PIR; A96695; A96695.
DR HSSP; P14779; 1JPZ.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00463; EP4501.
DR PRINTS; PR00385; P450.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 512 AA; 58463 MW; 038844B878935BC7 CRC64;

Query Match 16.4%; Score 78; DB 2; Length 512;
Best Local Similarity 31.1%; Pred. No. 50;
Matches 32; Conservative 13; Mismatches 36; Indels 22; Gaps 7;

QY 3 LLMVL--MLAALSOH-CYAGSGCPLLENVISTINPQVS--KTEYKELLQEFIDNNATTN 57
Db 227 LLTVLQRLCAQATRHLCFFGS-----RFLPSKYNREIKSLKTEVERLLMEIDSRKDSV 280

QY 58 AI-----DELKECFINQDETLSNVFVFMQLIYDSSLCDLF 93
Db 281 EIGRSSYGDLLGLLLNQMSNKNLNV--QMIMDE--CKTF 319

RESULT 22
Q6CK74 PRELIMINARY; PRT; 634 AA.
AC Q6CK74;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Kluyveromyces lacticis strain NRRL Y-1140 chromosome F of strain NRRL Y-1140 of Kluyveromyces lacticis.
GN ORFNames=KLIA0P130029;
OS Kluyveromyces lacticis NRRL Y-1140.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=284590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durkens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anhouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C., Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Niclaud J.M., Nikolski M., Ostas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zenciou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissbach J., Wincker P., Souciet J.L.;
RA "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382126; CAG98373.1; -.
SQ SEQUENCE 634 AA; 74302 MW; 6DDB2EAC7D8487B4 CRC64;

Query Match 16.2%; Score 77; DB 2; Length 634;
Best Local Similarity 33.8%; Pred. No. 78;
Matches 27; Conservative 10; Mismatches 29; Indels 14; Gaps 3;

QY 23 PLEENVISKTINPQVSKTEYKELLQEFIDNNATTNAIDELKECF-LNOTDETLSNVVEF- 80
Db 192 PTYKKKLNIILNKETETFYNNVLQEFIDLMKATLKIKGIQVCFVLNLT-----TNVEKEF 247

QY 81 -----MQLIYDSSLCDL 92
Db 248 NNLDSIIRALSXYDLKTIDL 267

RESULT 23
VAOD_SCHPO STANDARD; PRT; 343 AA.
AC O13753;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Vacuolar ATP synthase subunit d (EC 3.6.3.14) (V-ATPase d subunit) (Vacuolar proton pump d subunit) (V-ATPase 39 kDa subunit) (V-ATPase subunit M39).
GN Name=vma6; ORFNames=SPAC17A2.03c;
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgueros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B., Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe.";
RN Nature 415:871-880(2002).
CC -!- FUNCTION: Vacuolar ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells. The active enzyme consists of a catalytic V1 domain attached to an integral membrane V0 proton pore complex. This subunit is a non-integral membrane component of the membrane pore domain and is required for proper assembly of the V0 sector. Might be involved in the regulated assembly of V1 subunits onto the membrane sector or alternatively may prevent the passage of protons through V0 pores.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate + H(+) (Out).
CC -!- SUBUNIT: V-ATPase is an heteromultimeric enzyme composed of a peripheral catalytic V1 complex (components A to H) attached to an integral membrane V0 proton pore complex (components: a, c, c', c'', and d).
CC -!- SIMILARITY: Belongs to the V-ATPase V0D/AC39 subunit family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z99292; CAB16567.1; -;  
 DR PIR; T37804; T37804.03c; -;  
 DR GeneDB\_Spombe; SPAC17A2.03c; -;  
 DR InterPro; IPR002843; ATPynt AC39sub.  
 DR Pfam; PF01992; VATP-synt AC39; 1.  
 KW ATP synthesis; Hydrogen ion transport; Hydrolase.  
 SQ SEQUENCE 343 AA; 3323 MW; F8936922C7EBA691 CRC64;  
 Query Match 16.1%; Score 76.5; DB 1; Length 343;  
 Best Local Similarity 32.2%; Pred. No. 46;  
 Matches 29; Conservative 9; Mismatches 35; Indels 17; Gaps 5;  
 QY 10 AALSOHCYAG-SCPLLENVISTINPOVSKTY-----KELQEFDDNATTNAIDE 61  
 DB 23 ALLEQHIYNSLSCESLED-----FRLQSSDYGGLANQSKLTSSISAKATEKLLDE 77  
 QY 62 LKCEFLNQDTLSNVFVFMQLIYDSSLCD 91  
 DB 78 F-DLIRQADETLISK---FMDYITYAYMID 103  
 RESULT 24  
 Q7RHE7 PRELIMINARY; PRT; 2169 AA.  
 AC Q7RHE7  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE GVF domain, putative.  
 GN Name=PY04040;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XN1;  
 RX PubMed=12368865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perlea M.,  
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabib A., Cummings L.M.,  
 RA Florens L., Yates P.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 RT parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519(2002).  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AABL01001210; EAA15953.1; -;  
 DR InterPro; IPR011591; Botulinum.  
 DR InterPro; IPR003169; GVF.  
 DR Pfam; PF02213; GVF; 1.  
 DR ProDom; PD001963; Botulinum; 1.  
 DR PROSITE; PS00829; GVF; 1.  
 SQ SEQUENCE 2169 AA; 253886 MW; F48AD00EFA15E1CB CRC64;  
 Query Match 16.1%; Score 76.5; DB 2; Length 2169;  
 Best Local Similarity 36.4%; Pred. No. 3.1e+02;  
 Matches 20; Conservative 11; Mismatches 19; Indels 5; Gaps 2;  
 QY 38 SKTEYKELQEFDDNATTNAIDELKECFNLQDTLSNVFVFMQLIYDSSLCDL 92

DB 903 SKKKKNVILGE-INDSTNNVDQKRLINSEDSNFQN-----NQLINDNNMYSL 952  
 RESULT 25  
 UTER\_RABIT STANDARD; PRT; 91 AA.  
 ID UTER\_RABIT  
 AC P02779;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Uteroglobin precursor (Blastokinin).  
 GN Name=SCGB1A1; Synonym=UGB, UGL;  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83290960; PubMed=6309802;  
 RA Bailey A., Atger M., Atger P., Cerbon M.-A., Alizon M., Vu Hai M.T.,  
 RA Logeat F., Milgrom E.;  
 RT "The rabbit uteroglobin gene. Structure and interaction with the  
 RT progesterone receptor.";  
 RL J. Biol. Chem. 258:10384-10389(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83220783; PubMed=6304644;  
 RA Suske G., Wenz M., Cato A.C.B., Beato M.;  
 RT "The uteroglobin gene region: hormonal regulation, repetitive elements  
 RT and complete nucleotide sequence of the gene.";  
 RL Nucleic Acids Res. 11:2257-2271(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83014930; PubMed=6956897;  
 RA Menne C., Suske G., Arnenmann J., Wenz M., Cato A.C.B., Beato M.;  
 RT "Isolation and structure of the gene for the progesterone-inducible  
 RT protein uteroglobin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:4853-4857(1982).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83157105; PubMed=6299663;  
 RA Chandra T., Bullock D.W., Woo S.L.C.;  
 RT "Hormonally regulated mammalian gene expression: steady-state level  
 RT and nucleotide sequence of rabbit uteroglobin mRNA.";  
 RL DNA 1:19-26(1981).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82275176; PubMed=6287481;  
 RA Suske G., Menne C., Cato A., Wenz M., Beato M.;  
 RT "Characterization and sequence analysis of interspersed repetitive DNA  
 RT sequences transcribed in X.laavis embryos.";  
 RL Prog. Clin. Biol. Res. 85:139-146(1982).  
 RN [6]  
 RP SEQUENCE OF 1-73.  
 RX MEDLINE=79187160; PubMed=571719;  
 RA Atger M., Mercier J.-C., Haze G., Fridlansky F., Milgrom E.;  
 RT "N-terminal sequences of uteroglobin and its precursor.";  
 RL Biochem. J. 177:985-988(1979).  
 RN [7]  
 RP SEQUENCE OF 22-91.  
 RX MEDLINE=79042086; PubMed=568483;  
 RA Ponstingl H., Nieto A., Beato M.;  
 RT "Amino acid sequence of progesterone-induced rabbit uteroglobin.";  
 RL Biochemistry 17:3908-3912(1978).  
 RN [8]  
 RP SEQUENCE OF 22-91.  
 RX MEDLINE=79074850; PubMed=281700;  
 RA Popp R.A., Foresman K.R., Wise L.D., Daniel J.C. Jr.;  
 RT "Amino acid sequence of a progesterone-binding protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 75:5516-5519(1978).  
 RN [9]  
 RP REVISIONS TO 50-62 AND 67-71.



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:51:12 ; Search time 19.7582 Seconds  
(without alignments)  
452.882 Million cell updates/sec

Title: US-09-975-502a-5

Perfect score: 475

Sequence: 1 MKLLMVLMLAALSHQHCYAGS.....LSNVEFMQLIYDSSLCDLF 93

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR 79:\*

1: piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	204.5	43.1	95	1 BORT3	prostatic steroid-
2	86.5	18.2	95	2 S68231	PHG22 protein prec
3	78	16.4	512	2 A96895	hypothetical prote
4	76.5	16.1	343	2 T37804	vacuolar adenosine
5	76	16.0	91	1 UGRBL	uteroglobin precu
6	75	15.8	91	1 UGRBL	uteroglobin precu
7	74	15.6	1095	2 T13964	probable histone d
8	73.5	15.5	309	2 T09166	probable peroxidas
9	73	15.4	629	2 A96657	unknown protein, 7
10	73	15.4	1150	2 S49956	probable membrane
11	72.5	15.3	270	2 T48205	hypothetical prote
12	72.5	15.3	651	2 C75014	hypothetical prote
13	72	15.2	91	2 JS0036	Clara cell 10K pro
14	72	15.2	326	2 B71602	probable secreted
15	71.5	15.1	759	2 D71853	hypothetical prote
16	70	14.7	561	2 S71597	carboxylesterase (
17	69.5	14.6	270	2 AD1968	hypothetical prote
18	69.5	14.6	406	2 B97101	Zn-dependent pepti
19	69	14.5	144	1 JC4266	interleukin-3 prec
20	69	14.5	633	2 T28788	hypothetical prote
21	69	14.5	1021	2 B64576	hypothetical prote
22	68.5	14.4	446	2 H97754	lipid-A-disacchari
23	68.5	14.4	466	2 AH0476	NAD(P) transhydrog
24	68.5	14.4	764	2 H71607	hypothetical prote
25	68.5	14.4	785	2 D71653	cell surface antig
26	68	14.3	179	2 T48272	hypothetical prote
27	68	14.3	359	2 T44332	hypothetical prote
28	68	14.3	1127	2 T28317	ORF MSV156 hypote
29	67.5	14.2	724	2 G82437	SecA-related prote

30	67	14.1	146	1	I46407	interleukin-3 prec
31	67	14.1	265	2	T46173	hypothetical prote
32	67	14.1	368	2	C90558	lipoprotein (impor
33	67	14.1	449	2	F84241	hypothetical prote
34	67	14.1	679	2	S06000	penicillin-binding
35	66.5	14.0	282	2	S74617	prohibitin phb - S
36	66.5	14.0	323	2	T09164	probable peroxidas
37	66.5	14.0	359	2	S14028	type II site-speci
38	66.5	14.0	680	2	G95194	penicillin-binding
39	66.5	14.0	685	2	D98061	penicillinbinding p
40	66.5	14.0	688	2	S45803	hypothetical prote
41	66.5	14.0	1048	2	C96669	protein F1N19.15 [
42	66.5	14.0	1764	2	S37827	hypothetical prote
43	66	13.9	430	2	AE1803	sugar binding prot
44	66	13.9	430	2	AF1429	sugar binding prot
45	66	13.9	440	2	JS0374	hypothetical 51.6K
46	66	13.9	590	2	S16411	terminase Arpase c
47	66	13.9	618	2	AD1161	two-component sens
48	66	13.9	1558	2	B71603	RESA-H3 antigen PF
49	66	13.9	4859	2	S74173	ryanodine receptor
50	65.5	13.8	376	2	AD1621	heat shock protein
51	65.5	13.8	377	2	T43739	heat shock protein
52	65.5	13.8	377	2	AI1258	heat shock protein
53	65.5	13.8	649	1	W1WL11	B1 protein - human
54	65.5	13.8	698	2	D82538	polysphosphate kina
55	65.5	13.8	1071	1	PXBIVA	H+-exporting ATPas
56	65.5	13.8	4868	2	B54161	ryanodine-binding
57	65	13.7	1277	2	E70224	hypothetical prote
58	65	13.7	4872	2	S27272	ryanodine receptor
59	64.5	13.6	214	2	B72377	transcription regu
60	64.5	13.6	329	2	T04709	peroxidase (EC 1.1
61	64.5	13.6	623	2	AH1209	B. subtilis minor
62	64.5	13.6	830	2	C69011	conserved hypotet
63	64.5	13.6	1070	2	T00767	hypothetical prote
64	64.5	13.6	1137	2	S13759	morphogenesis-rela
65	64.5	13.6	1173	2	T52575	gigantea protein [
66	64.5	13.6	1967	2	S64604	hypothetical prote
67	64	13.5	96	2	A36581	polychlorinated bi
68	64	13.5	181	2	G86786	hypothetical prote
69	64	13.5	545	2	T12765	site-specific DNA
70	64	13.5	558	2	C96933	2-isopropylmalate
71	64	13.5	588	2	AH1037	probable terminase
72	64	13.5	3724	2	T18427	hypothetical prote
73	63.5	13.4	192	2	AG1138	hypothetical prote
74	63.5	13.4	424	2	T41449	probable phd finge
75	63.5	13.4	594	2	T04783	hypothetical prote
76	63.5	13.4	653	2	H96630	hypothetical prote
77	63.5	13.4	658	2	AI1385	exonuclease ABC (
78	63.5	13.4	720	2	A89928	hypothetical prote
79	63.5	13.4	1011	2	T17430	col protein - Neur
80	63.5	13.4	1087	2	T30330	gelsolin-related p
81	63	13.3	111	1	BORT1	prostatic steroid-
82	63	13.3	266	2	A82851	chaperone protein-
83	63	13.3	364	2	T39235	probable translati
84	63	13.3	411	2	A44121	ribosomal protein
85	63	13.3	435	2	S69035	hypothetical prote
86	63	13.3	476	2	A44170	membrane-bound rib
87	63	13.3	529	2	T46130	RNA polymerase III
88	63	13.3	554	2	B82334	hypothetical prote
89	63	13.3	561	2	S62788	carboxylesterase (
90	63	13.3	588	2	AB0927	terminase, Arpase
91	63	13.3	673	2	H72563	probable DNA topoi
92	63	13.3	1829	2	T26135	hypothetical prote
93	62.5	13.2	187	2	AH2190	polypeptide deform
94	62.5	13.2	344	2	D97761	hypothetical prote
95	62.5	13.2	349	1	E64442	probable arsenical
96	62.5	13.2	412	2	T24441	hypothetical prote
97	62.5	13.2	430	2	AH1248	probable proteinas
98	62.5	13.2	612	2	E81287	probable sugar tra
99	62.5	13.2	625	2	F81287	probable sugar tra
100	62.5	13.2	658	2	AC1761	exonuclease ABC (

ALIGNMENTS

RESULT 1  
BORT3  
N;Alternate names: prostacin  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 17-Dec-1982 #sequence\_revision 14-Nov-1983 #text\_change 09-Jul-2004  
C;Accession: A92395; A92433; A91108; A42392; A03250  
R;Parker, M.G.; White, R.; Hurst, H.; Needham, M.; Tilly, R.  
J. Biol. Chem. 258, 12-15, 1983  
A;Title: Prostatic steroid-binding protein. Isolation and characterization of C3 genes.  
A;Reference number: A92395; MUID:83082848; PMID:6294095  
A;Accession: A92395  
A;Molecule type: mRNA  
A;Residues: 1-95 <PAR>  
A;Cross-references: UNIPROT:P02780; GB:V01263; GB:J00777; NID:956993; PIDN:CAA24577.1; F  
R;Viskochil, D.H.; Perry, S.T.; Lea, O.A.; Stafford, D.W.; Wilson, E.M.; French, F.S.  
J. Biol. Chem. 258, 8861-8866, 1983  
A;Title: Isolation of two genomic sequences encoding the M-r = 14,000 subunit of rat pro  
A;Reference number: A92433; MUID:83238526; PMID:6190812  
A;Accession: A92433  
A;Molecule type: mRNA  
A;Residues: 1-78, 'S', 80-95 <VIS>  
R;Feetters, B.; Rombauts, W.; Mous, J.; Heyns, W.  
Eur. J. Biochem. 115, 115-121, 1981  
A;Title: Structural studies on rat prostatic binding protein. The primary structure of i  
A;Reference number: A91108; MUID:81188769; PMID:7014218  
A;Accession: A91108  
A;Molecule type: protein  
A;Residues: 19-95 <PEE>  
R;Tan, J.A.; Marschke, K.B.; Ho, K.C.; Perry, S.T.; Wilson, E.M.; French, F.S.  
J. Biol. Chem. 267, 4456-4466, 1992  
A;Title: Response elements of the androgen-regulated C3 gene.  
A;Reference number: A42392; MUID:92165796; PMID:1537831  
A;Accession: A42392  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-16, 'T', 18-95 <TAN>  
A;Note: sequence inconsistent with the nucleotide translation  
A;Note: sequence extracted from NCBI backbone (NCBIN:83851, NCBIIP:83856)  
C;Comment: C3 is encoded by two unique genes that differ from each other only in their i  
ng C2 and C3 chains. The chains of each dimer are linked by disulfide bonds. This protei  
C;Superfamily: uteroglobin  
C;Keywords: glycoprotein  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-95/Product: prostatic steroid-binding protein C3 chain #status experimental <MAT>  
F;35/Binding site: carbohydrate (Asn) #status experimental

Query Match 43.1%; Score 204.5; DB 1; Length 95;  
Best Local Similarity 42.9%; Pred. No. 8.6e-13;  
Matches 39; Conservative 22; Mismatches 29; Indels 1; Gaps 1;  
Query 1 MKLLVLMIAALSOHCYA-GSGCPLLENVISKINPQVSKTEYKELLOFIDNATTNAI 59  
Db 1 MKLVFLFLVLTPICCYASGSGSILDEVIRGTINSTVLHDYMKLVKPYVDHTEKAV 60  
Query 60 DELKECFLNQTDETLSNVFVMQIYDSSLIC 90  
Db 61 KQPKQCFLDQTDKTLNVGVGMEAFNFSSEC 91

RESULT 2  
S68231  
FHG22 protein precursor - golden hamster  
C;Species: Mesocricetus auratus (golden hamster)  
C;Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 21-Aug-1998  
C;Accession: S68231  
R;Domínguez, P.  
FEBS Lett. 376, 257-261, 1995

A;Title: Cloning of a Syrian hamster cDNA related to sexual dimorphism: establishment of  
A;Reference number: S68231; MUID:96105393; PMID:7498554  
A;Accession: S68231  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-95 <DOM>  
A;Cross-references: EMBL:Z66540  
C;Superfamily: uteroglobin  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-95/Product: FHG22 protein #status predicted <MAT>

Query Match 18.2%; Score 86.5; DB 2; Length 95;  
Best Local Similarity 33.3%; Pred. No. 0.19;  
Matches 28; Conservative 15; Mismatches 26; Indels 15; Gaps 5;  
Query 1 MKLLVLMIAALSOHCY---AGSGCPLLENVISKINPQVSKTEYKELLOFIDNATTN 57  
Db 1 MKLSCLLLVILAVHCYEAANAVVCPAVLSVSKSFLFDKVEK--FEAYLQTF---NAPPE 55  
Query 58 AID---ELKECFLNQTDETLSNVE 78  
Db 56 AVKAKVEVKCI----DSTLVNLE 75

RESULT 3  
A96695  
hypothetical protein F5A8.3 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: A96695  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;  
C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.  
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: A96695  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-512 <STO>  
A;Cross-references: UNIPROT:Q9ZW95; GB:AE005173; NID:94204278; PIDN:AAID0659.1; GSPDB:GN  
C;Genetics:  
A;Gene: F5A8.3  
A;Map position: 1  
C;Superfamily: human cytochrome P450 CYP4B1; cytochrome P450 homology  
C;Keywords: heme; iron; metalloprotein  
F;458/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 16.4%; Score 78; DB 2; Length 512;  
Best Local Similarity 31.1%; Pred. No. 7.7;  
Matches 32; Conservative 13; Mismatches 36; Indels 22; Gaps 7;  
Query 3 LLWL--MLAALSOH-CYAGSGCPLLENVISKINPQVS--KTEYKELLOFIDNATTN 57  
Db 227 LLTVLQRLCAQATRHLCFFGS-----RFLPSKYNREIKSLKTEVERLLMEIDSRKDSV 280  
Query 58 AI-----DELKECFLNQTDETLSNVFVMQIYDSSLICLF 93  
Db 281 EIGRSSSYGDDLLGLLNQMDSNKNLNV--QMIND--CKTF 319

RESULT 4  
T37804  
vacuolar adenosine triphosphatase subunit d - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37804  
R;Olliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A/Reference number: Z21748

A/Accession: T37804

A/Status: preliminary; translated from GB/EMBL/DBJ/

A/Molecule type: DNA

A/Residues: 1-343 <OLI>

A/Cross-references: UNIPROT:O13753; EMBL:Z99292; P

A/Experimental source: strain 972h-; cosmid c17A2

C/Genetics:

A/Gene: SPDB:SPAC17A2.03c

A/Map position: 1

A/Introns: 7/2

```

Query Match      16.1%; Score 76.5; DB 2; Length 343;
Best Local Similarity 32.2%; Pred. No. 6.9;
Matches 29; Conservative 9; Mismatches 35; Indels 17; Gaps 5;

Qy      10 AALSOHCYAG-SGCPLELNVISKINPOVSKTEY-----KELQEFIDDNATNAIDE 61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      23 ALLEQHIYSNLSQCESLED-----FRQLQSTDYGGFLANQSKLTSSIIASAKATEKLIDE 77

Qy      62 LKECFNLQTDETLSNVVEVFMQLIYDSSLCD 91
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      78 F-DLIRROADETLSK---FMDIYIYAMID 103

```

## RESULT 5

uteroglobulin precursor [validated] - rabbit  
N;Alternate names: blastokinin  
C;Species: Oryctolagus cuniculus (domestic rabbit)  
C;Date: 28-Feb-1980 #sequence revision 15-Oct-1982 #text change 09-Jul-2004  
A;Accession: A92391; A93461; A90935; A24217; A90303; A9017; A93824; A94608; I46904; I46904; F.; R.; Bailly, A.; Atger, M.; Aguer, P.; Carbon, M.A.; Allizon, M.; Vu Hai, M.T.; Logeat, F.; J. Biol. Chem. 258, 10384-10389, 1983  
A;Title: The rabbit uteroglobin gene. Structure and interaction with the progesterone receptor.  
A;Reference number: A92391; PMID:83290960; PMID:6309802  
A;Accession: A92391  
A;Molecule type: DNA  
A;Residues: 1-91 <BAI>  
A;Cross-references: UNIPROT:P02779; GB:X00049; NID:g165789  
R;Suske, G.; Wenz, M.; Cato, A.C.B.; Beato, M.  
Nucleic Acids Res. 11, 2257-2271, 1983  
A;Title: The uteroglobin gene region: hormonal regulation, repetitive elements and complexed structures.  
A;Reference number: A93461; PMID:83220783; PMID:6304644  
A;Accession: A93461  
A;Molecule type: DNA  
A;Residues: 1-91 <SUS>  
A;Cross-references: GB:J00687; NID:g1772; PIDN:CAA25669.1; PID:g313668  
R;Chandra, T.; Bullock, D.W.; Woo, S.L.C.  
DNA 1, 19-26, 1981  
A;Title: Horizontally regulated mammalian gene expression: steady-state level and nucleotide sequence analysis.  
A;Reference number: A90935; PMID:83157105; PMID:6299663  
A;Accession: A90935  
A;Molecule type: mRNA  
A;Residues: 1-91 <CHA>  
A;Cross-references: GB:K01657; NID:g165794; PIDN:AAA31497.1; PID:g165795  
R;Lopez de Haro, M.S.; Nieto, A.  
FEBS Lett. 193, 247-249, 1985  
A;Title: Primary structure of rabbit lung uterogloblin as deduced from the nucleotide sequence.  
A;Reference number: A24217; PMID:86056319; PMID:2415398  
A;Accession: A24217  
A;Molecule type: mRNA  
A;Residues: 22-91 <LOP>  
A;Cross-references: GB:M27564; NID:g165792; PIDN:AAA31496.1; PID:g165793  
A;Experimental source: lung  
R;Atger, M.; Mercier, J.C.; Haze, G.; Fridlansky, F.; Milgrom, E.  
Biochem. J. 177, 985-988, 1979  
A;Title: N-terminal sequences of uterogloblin and its precursor.  
A;Reference number: A90303; PMID:79187160; PMID:571719  
A;Accession: A90303  
A;Molecule type: protein  
A;Residues: 1-5,'F','7'-10,  
X',15,'G',17-54,'X',56,'B',58-66,'B',68-70,'XX',73 <ATG>

R;Ponstingl, H.; Nieto, A.; Beato, M.  
Biochemistry 17, 3908-3912, 1978  
A:Title: Amino acid sequence of progesterone-induced rabbit uteroglobin.  
A:Reference number: A90417; MUID:79042086; PMID:568483  
A:Accession: A90417  
A:Molecule type: protein  
A:Residues: 22-81,'Q',83-91 <PON>  
R;Popp, R.A.; Foresman, K.R.; Wise, L.D.; Daniel Jr., J.C.  
Proc. Natl. Acad. Sci. U.S.A. 75, 5516-5519, 1978  
A:Title: Amino acid sequence of a progesterone-binding protein.  
A:Reference number: A93824; MUID:79074850; PMID:281700  
A:Accession: A93824  
A:Molecule type: protein  
A:Residues: 22-49,'D',51,'EN',54-59,61-66,'NEPSL',72-91 <POP>  
R;Popp, R.A.; Foresman, K.R.; Wise, L.D.; Daniel Jr., J.C.  
submitted to the Atlas, October 1982  
A:Reference number: A94608  
A:Accession: A94608  
A:Molecule type: protein  
A:Residues: 50-62,167-71 <PO2>  
R;Morize, I.; Surcouf, E.; Vaney, M.C.; Buehner, M.; Mornon, J.P.  
submitted to the Brookhaven Protein Data Bank, April 1989  
A:Reference number: A50025; PDB:1UTG  
A:Contents: annotation: X-ray crystallography, 1.34 angstroms, residues 22-91  
R;Morize, I.; Surcouf, E.; Vaney, M.C.; Epelboin, Y.; Buehner, M.; Fridlansky, F.; Milgrom, J.  
J. Mol. Biol. 194, 725-739, 1987  
A:Title: Refinement of the C222-1 crystal form of oxidized uteroglobin at 1.34 angstroms  
A:Reference number: A4452; PMID:88011213; PMID:3656405  
A:Contents: annotation: X-ray crystallography, 1.34 angstroms  
R;Bally, R.; Delettre, J.  
submitted to the Brookhaven Protein Data Bank, May 1989  
A:Reference number: A50553; PDB:2UTG  
A:Contents: annotation: X-ray crystallography, 1.64 angstroms, residues 22-91  
R;Bally, R.; Delettre, J.  
J. Mol. Biol. 206, 153-170, 1989  
A:Title: Structure and refinement of the oxidized P2-1 form of uteroglobin at 1.64 angstroms  
A:Reference number: A4453; MUID:8919637; PMID:2704039  
R;Contents: annotation: X-ray crystallography, 1.64 angstroms; disulfide bonds  
R;Menne, C.; Suske, G.; Arneemann, J.; Wenz, M.; Cato, A.C.B.; Beato, M.  
Proc. Natl. Acad. Sci. U.S.A. 79, 4853-4857, 1982  
A:Title: Isolation and structure of the gene for the progesterone-inducible protein uteroglobin  
A:Reference number: I46904; MUID:83014990; PMID:6956897  
A:Accession: I46904  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-91 <MEN>  
A:Cross-references: GB:J00689; NID:g165786; PIDN:AAA31495.1; PID:g165788  
R;Chandra, T.; Woo, S.L.C.; Bullock, D.W.  
Biochem. Biophys. Res. Commun. 95, 197-204, 1980  
A:Title: Cloning of the rabbit uteroglobin structural gene.  
A:Reference number: I46905; MUID:81021016; PMID:7417250  
A:Accession: I46905  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 39-77 <CH2>  
A:Cross-references: GB:M25057; NID:g165802; PIDN:AAA31498.1; PID:g165803  
R;Suske, G.; Menne, C.; Cato, A.; Wenz, M.; Beato, M.  
Proc. Clin. Biol. Res. 85, 139-146, 1982  
A:Title: Structure and regulated expression of the uteroglobin gene.  
A:Reference number: I46907; MUID:82275176; PMID:6287481  
A:Accession: I46907  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-45,'V',47-91 <SU2>  
A:Cross-references: GB:M20212; NID:g165807; PIDN:AAA31500.1; PID:g165809  
R;Atger, M.; Perricaudet, M.; Tiollais, P.; Milgrom, E.  
Biochem. Biophys. Res. Commun. 93, 1082-1088, 1980  
A:Title: Bacterial cloning of the rabbit uteroglobin structural gene.  
A:Reference number: I46906; MUID:80241888; PMID:6156676  
A:Accession: I46906  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 53-66,'NT',69-72 <AT2>



A;Cross-references: GB:M25038; NID:g165804; PIDN:AAA31499.1; PID:g165805  
C;Comment: Utteroglobin is secreted by the uterus upon induction by progesterone. It binds  
C;Genetics:  
A;Introns: 19/1; 81/3  
C;Complex: homodimer linked by two disulfide bonds  
C;Superfamily: uteroglobin  
C;Keywords: homodimer; steroid binding; uterus  
F;1-21/Domain: signal sequence #status experimental <SIG>  
F;22-91/Product: uteroglobin #status experimental <MAT>  
F;24/Disulfide bonds: interchain (to 90) #status experimental  
F;90/Disulfide bonds: interchain (to 24) #status experimental

Query Match 16.8%; Score 76; DB 1; Length 91;  
Best Local Similarity 27.1%; Pred.No. 1.8;  
Matches 26; Conservative 17; Mismatches 41; Indels 12; Gaps 3;

QY 1 MKLLMVLMLAALSOHCYAGSG--CP-----LLENVISKTNIPQVSKTEYKELLQEFIDDNA 54  
||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 1 MKLAITLAVTLALLCSPASGICPRFAHVIENLLGT-----PSSYETSLKEFPDDT 54  
||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | :

QY 55 TTNAIDELKECFNLQNTDTELNSVEVFQMQLIYDSLC 90  
: ||| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 55 MKDAGMQMKVLDLPQTTRINIKLTEKIVKSPCLC 90  
: ||| : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 6  
UGREL  
uteroglobin precursor - brown hare  
N;Alternate names: blastokinin  
C;Species: Lepus capensis (brown hare)  
C;Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jul-2004  
C;Accession: A23825  
R;Lopez de Haro, M.S.; Nieto, A.  
Biochem. J. 235, 895-898, 1986  
A;Title: Nucleotide and derived amino acid sequences of a cDNA coding for pre-uterogloblin  
A;Reference number: A23825; MUID:86323069; PMID:3019311  
A;Accession: A23825  
A;Molecule type: mRNA  
A;Residues: 1-91 <LOP>  
A;Cross-references: UNIPROT:P06913; GB:M25609; NID:g164246; PIDN:AAA30960.1; PID:g164247  
A;Experimental source: lung  
C;Comment: Uteroglobin, synthesized in the uterus and lung, is secreted by the uterus up  
C;Complex: homodimer linked by two disulfide bonds  
C;Superfamily: uteroglobin  
C;Keywords: lung; steroid binding; uterus  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-91/Product: uteroglobin #status predicted <MAT>  
F;24/Disulfide bonds: interchain (to 90) #status predicted  
F;90/Disulfide bonds: interchain (to 24) #status predicted

Query Match 15.8%; Score 75; DB 1; Length 91;  
Best Local Similarity 28.1%; Pred.No. 2.3;  
Matches 27; Conservative 15; Mismatches 42; Indels 12; Gaps 3;

QY 1 MKLLMVLMLAALSOHCYAGSG--CP-----LLENVISKTNIPQVSKTEYKELLQEFIDDNA 54  
||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 1 MKLTITLAVTLALLCSPASGICPGFAHVIENLLGT-----PSSYGTSLKFEFPDDA 54  
||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | :

QY 55 TTNAIDELKECFNLQNTDTELNSVEVFQMQLIYDSLC 90  
: ||| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 55 MKDAGMQMKVLDLPQTTRINIKLTEKIVKSPCLC 90  
: ||| : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 7  
T13964  
Probable histone deacetylase (EC 3.5.1.-) HDA2 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 12-Nov-1999  
C;Accession: T13964  
R;Verdel, A.; Khochbin, S.  
J. Biol. Chem. 274, 2440-2445, 1999  
A;Title: Identification of a new family of higher eukaryotic histone deacetylases. Coord  
A;Reference number: Z17841; MUID:99107904; PMID:9891014



A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: A96657  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-629 <STO>  
A;Cross-references: UNIPROT:O9CAM8; GB:AE005173; NID:96598835; PIDN:AAF18690.1; GSPDB:GN

C;Genetics:  
A;Gene: F1GM19.13  
A;Map position: 1  
Query Match 15.4%; Score 73; DB 2; Length 629;  
Best Local Similarity 23.7%; Pred. No. 29;  
Matches 23; Conservative 16; Mismatches 28; Indels 30; Gaps 3;

Qy 24 LLENVSKTINPQV-----SKTEYKELQELPIDNATT----- 56  
Db 315 LLSNMLEKKINPNVVTNALIDAFFKEGKLVAEKLEHMIQRSIDPTIYNLLINGFC 374  
Qy 57 --NAIDELKECF-LNQDTETLSNVVEFMQLIYDSSL 90  
Db 375 MHNRLDEAKQMFPMVSKDCLPNITQYNTLINGFCKC 411

RESULT 10  
S49956  
probable membrane protein YIL026c - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: hypothetical protein YI3299.05C  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 09-Jul-2004  
A;Accession: S49956; S50263; S57371  
R;Skellton, J.; Churcher, C.  
submitted to the EMBL Data Library, December 1994  
A;Reference number: S49951  
A;Accession: S49956  
A;Molecule type: DNA  
A;Residues: 1-1150 <SKE>  
A;Cross-references: UNIPROT:P40541; EMBL:Z46881; NID:G599967; PID:G599973; MIPS:YIL026c  
R;Kurlandzka, A.; Gromadka, R.; Murawski, M.  
submitted to the EMBL Data Library, December 1994  
A;Description: A new essential gene located on *Saccharomyces cerevisiae* chromosome IX.  
A;Reference number: S50263  
A;Accession: S50263  
A;Molecule type: DNA  
A;Residues: 1-938, 'G', 940-1072, 'TYRKQNIQRKGRKLYNTVNEKPTMQLIKIMIQLFP' <KUR>  
A;Cross-references: EMBL:U17918  
R;Kurlandzka, A.; Rycka, J.; Gromadka, R.; Murawski, M.  
Yeast 11, 885-890, 1995  
A;Title: A new essential gene located on *Saccharomyces cerevisiae* chromosome IX.  
A;Reference number: S57371; MUID:96090137; PMID:7483852  
A;Accession: S57371  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-938, 'G', 940-1150 <KUR>  
A;Cross-references: EMBL:U17918; NID:G619594; PIDN:AAAC49039.1; PID:G619595

C;Genetics:  
A;Gene: SGD:IRR1  
A;Cross-references: SGD:S0001288; MIPS:YIL026c  
A;Map position: 9L  
C;Keywords: transmembrane protein  
F;546-562/Domain: transmembrane #status predicted <TM1>  
F;896-912/Domain: transmembrane #status predicted <TM2>  
F;940-946/Domain: transmembrane #status predicted <TM3>  
Query Match 15.4%; Score 73; DB 2; Length 1150;  
Best Local Similarity 25.3%; Pred. No. 56;  
Matches 22; Conservative 19; Mismatches 20; Indels 26; Gaps 3;

Qy 22 CPLENVISKTNPQVS---KTEYKEL-----LQEFIDNATTNAID----- 60  
Db 700 CPLESPTTSLDDWISIGVETELKKTIDTKIFAMDSIGNSKVDKMYDIFAKFIHHI 759

Qy 61 -----ELKECFLNQDTETLSNVVEFMQ 82  
Db 760 HFEKSELQELFNQIATLKHKLKFLQ 786

## RESULT 11

T48205  
hypothetical protein T20L15.110 - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C;Accession: T48205

R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkee, W.; Stiekema, W.; Bancroft, I.; Mew  
submitted to the Protein Sequence Database, March 2000  
A;Reference number: Z24488  
A;Accession: T48205  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-270 <BEV>  
A;Cross-references: UNIPROT:Q9LZW2; EMBL:AL162351  
A;Experimental source: cultivar Columbia; BAC clone T20L15  
C;Genetics:  
A;Map position: 5  
A;Note: T20L15.110

Query Match 15.3%; Score 72.5; DB 2; Length 270;  
Best Local Similarity 27.6%; Pred. No. 13;  
Matches 21; Conservative 17; Mismatches 31; Indels 7; Gaps 3;

Qy 18 AGSGCPLENVISKTNPQVSKTEYKELQELTDDN--ATTNAIDELKECF--LNQTDET 73  
Db 192 SGSSARRRAVVASVDP---KRDFKSEEMTAENKIRATKOLELLACYLCLNSDEYH 248

Qy 74 LSNVVEFMQLIYDSSL 89  
Db 249 AIIINVFKQIWLNL 264

## RESULT 12

C75014  
hypothetical protein PAB1258 - *Pyrococcus abyssi* (strain Orsay)

C;Species: *Pyrococcus abyssi*  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: C75014

R;anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure  
A;Reference number: A75001

A;Accession: C75014

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-651 <KAW>

A;Cross-references: UNIPROT:Q9UY59; GB:AJ248288; NID:G5458960; PIDN:CAB5055

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB1258

C;Superfamily: *Pyrococcus abyssi* hypothetical protein PAB1258

Query Match 15.3%; Score 72.5; DB 2; Length 651;  
Best Local Similarity 31.3%; Pred. No. 34;  
Matches 31; Conservative 12; Mismatches 29; Indels 27; Gaps 5;

Qy 7 LMLAALSQHCYAGSG-----CPLLE---NVISKTNPQVSKTE---YKELLQ 47  
Db 464 LLATMLQQATITGSELVEVARKISEIQLELRNRGVIRKAVNKIVSSDIPPGYKELLR 523

Qy 48 EFIDDNATTNAIDELKECFLNQDTETLSNVVEFMQLIYD 86  
Db 524 E-IEDPMLRKILERKE-----NVIRSVAVYKLLQD 554

## RESULT 13

JS0036



F;221,466/Active site: Ser, His #status predicted

Query Match 14.7%; Score 70; DB 2; Length 561;  
Best Local Similarity 27.8%; Pred. No. 50;  
Matches 25; Conservative 17; Mismatches 36; Indels 12; Gaps 3;  
QY 6 VLMALASQHCY----AGSCPLLENVSKTINPOVSKTEYKELLOEFIDNATTAIDE 61  
DB 229 VLVLSPUSKNIYHRAISESGVLIQTQKDVPAKQ-----IADMAGCKTTTSAI-- 280

QY 62 LKECFLNQDTETLSNVVEFMQLIYDSSLCD 91  
DB 281 IVHCLRQKTEBELLEMEKNLIKLSQORD 310

## RESULT 17

AD1968  
hypothetical protein alr1295 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C;Accession: AD1968  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AD1968  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-270 <KUR>  
A;Cross-references: UNIPROT:Q8YXC0; GB:BA000019; PIDN:BA073252.1; PID:g17130642; GSPDB:G000000000  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr1295  
C;Superfamily: erythrocyte band 7 integral membrane protein

Query Match 14.6%; Score 69.5; DB 2; Length 270;  
Best Local Similarity 29.4%; Pred. No. 25;  
Matches 20; Conservative 14; Mismatches 29; Indels 5; Gaps 2;  
QY 25 LENVSKTINPOVSKTEYKELLOEFIDNATTAIDELEKCEFLNQDTETLSNVVEFMQLI 84  
DB 102 LENIVSKIIAPQEQEAPKIAAARTVEAITKRS--ELKEDFONALGRLDKYGI---IV 156

QY 85 YDGSGLCDL 92  
DB 157 LDTSVVDL 164

## RESULT 18

B97101  
Zn-dependent peptidase from MPP family [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: B97101  
R;Nolling, J.; Bretton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: B97101  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-406 <KUR>  
A;Cross-references: UNIPROT:Q97ILO; GB:AE001437; PIDN:AAK79597.1; PID:g15024588; GSPDB:G000000000  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC1630

Query Match 14.6%; Score 69.5; DB 2; Length 406;  
Best Local Similarity 25.6%; Pred. No. 39;  
Matches 23; Conservative 14; Mismatches 30; Indels 23; Gaps 4;

QY 21 GCPLLENVSKTINPOVSK-----TEYKELLOEFIDNATTAID--ELKE 64  
DB 96 GFELYSDII---VNPTFSEGFEEKSIICEELTEWKDDKQCFCEDELLKNSFSNIRLKE 152  
QY 65 CFL-----NOTDETLSNVVEFMQLIYDSSLC 90  
DB 153 CIIIGNEKNIKPSIDELRKYKYKYYTSDNC 182

## RESULT 19

JC4266  
interleukin-3 precursor - bovine  
N;Alternate names: hematopoietic growth factor; IL-3; mast-cell growth factor; multi-CSF  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 15-Oct-1995 #sequence\_revision 15-Nov-1996 #text\_change 09-Jul-2004  
C;Accession: JC4266  
R;Mwangi, S.M.; Logan-Henfrey, L.; McInnes, C.; Mertens, B.  
Gene 162, 309-312, 1995  
A;Title: Cloning of the bovine interleukin-3-encoding cDNA.  
A;Reference number: JC4266; MUID:96032363; PMID:7557449  
A;Accession: JC4266  
A;Molecule type: mRNA  
A;Residues: 1-144 <MWA>  
A;Cross-references: UNIPROT:P49875; GB:L31893; NID:G473414; PIDN:AAA99502.1; PID:G473415  
C;Comment: This cytokine, produced by bone marrow cells, by mitogen or antigen-activated  
tiation of Hematopoietic cells.  
C;Superfamily: interleukin-3  
C;Keywords: cytokine; growth factor; lymphokine; mitogen; monomer; T-cell  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;21-144/Product: interleukin-3 #status predicted <MAT>

Query Match 14.5%; Score 69; DB 1; Length 144;  
Best Local Similarity 22.5%; Pred. No. 14;  
Matches 20; Conservative 21; Mismatches 32; Indels 16; Gaps 2;

QY 4 LMVLMALASQHCYAGSGCPLLENVSKTINPOVSKTEYKELLOEFIDD----NATTNA 58  
DB 6 ILHLLLLALHAPOAKGUPV-----TSRTPYMLMKKEIMDDLKKITPSPEGS 54  
QY 59 IDELKECFNLQDTETLSNVVEFMQLIYDS 87  
DB 55 LNSDEKNFLTKESLLQANLKVFTFATDT 83

## RESULT 20

T28788  
hypothetical protein C41D11.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T28788  
R;Gattung, S.; Maggi, L.  
submitted to the EMBL Data Library, May 1997  
A;Description: The sequence of C. elegans cosmid C41D11.  
A;Reference number: Z20522  
A;Accession: T28788  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-633 <GAT>  
A;Cross-references: UNIPROT:O01976; EMBL:AF003740; PIDN:AAC48142.1; GSPDB:GN000019; CBSP:  
A;Experimental source: strain Bristol N2; clone C41D11  
C;Genetics:  
A;Gene: CBSP:C41D11.7  
A;Map position: 1  
A;Introns: 22/2; 153/1; 177/3; 211/3; 283/2; 344/2; 402/1; 483/1; 527/2  
C;Superfamily: probable DNA helicase MJ0104

Query Match 14.5%; Score 69; DB 2; Length 633;  
Best Local Similarity 23.2%; Pred. No. 71;  
Matches 19; Conservative 19; Mismatches 26; Indels 18; Gaps 3;

QY 27 NVLSKTNPOVSKTEYKELLOEFIDNATTAIDELKECFNLQ-----DETLSNVEV-- 79

[illegible]

R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A;Reference number: A71630; MUID:99039499; PMID:9823893  
A;Accession: D71653  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-785 <AND>  
A;Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PIDN:CAA14950.1; PID:el34279  
A;Experimental source: strain Madrid E  
C;Genetics:  
A;Gene: sca4; RP498

Query Match 14.4%; Score 68.5; DB 2; Length 785;  
Best Local Similarity 26.4%; Pred. No. 1e+02;  
Matches 19; Conservative 17; Mismatches 27; Indels 9; Gaps 3;  
Qy 24 LLENVSKTINPQVSKTEYKELLQEFIDNNATT-----NAIDELKECFLNQTDET---LS 75  
Db 351 LLDSILKKTAEQTQLSPQQQLMNQL-DNITTEHTKRDTEKVNILLEPLSNTALKTT 409  
Qy 76 NVEFMOLIYDS 87  
Db 410 NIQVMTSNVLDS 421

Search completed: September 26, 2005, 08:31:07  
Job time : 24.7582 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:18:37 ; Search time 102.857 Seconds  
(without alignments)  
338.415 Million cell updates/sec

Title: US-09-975-502A-6

Perfect score: 450

Sequence: 1 MKLSVCLLVLTALCCYQAN.....LQKRSLIAEVLVKILKCSV 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A\_Geneseq\_16Dec04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	450	100.0	90	2 AAW35803	Raw35803 Human end
2	450	100.0	90	2 AAW54271	Aaw54271 BUI01 ant
3	450	100.0	90	2 AAW59776	Aaw59776 Amino aci
4	450	100.0	90	2 AAW89613	Raw89613 Endometri
5	450	100.0	90	3 AAY84875	Aay84875 A human e
6	450	100.0	90	3 AAB03768	Abp03768 Human end
7	450	100.0	90	3 AAB13787	Abp13787 Human BUI
8	450	100.0	90	3 AAB07501	Abp07501 Amino aci
9	450	100.0	90	4 AAG65989	Rag65989 Lipophilli
10	450	100.0	90	4 AAB31681	Rag31681 An endome
11	450	100.0	90	4 AAE07518	Rae07518 Human lip
12	450	100.0	90	5 ABG96363	Abp96363 Human ova
13	450	100.0	90	5 ABB09634	Abp09634 Human end
14	450	100.0	90	5 ABJ05577	Abj05577 Breast ca
15	450	100.0	90	6 ABJ19810	Abj19810 Androgen-
16	450	100.0	90	6 ABP96089	Abp96089 Human lip
17	450	100.0	90	6 ABP96097	Abp96097 Human lip
18	450	100.0	90	6 ABG73168	Abg73168 Human end
19	450	100.0	90	6 ABR47518	Abp47518 Breast ca
20	450	100.0	90	6 ADA43319	Ada43319 Human BUI
21	450	100.0	90	7 AAO24005	Aao24005 Human BUI
22	450	100.0	90	8 ADN04265	Adn04265 Antipsori
23	450	100.0	90	8 ADQ19742	Adq19742 Human sof
24	450	100.0	90	8 ADR46900	Adr46900 Breast ca
25	450	100.0	117	4 ABB11907	Abb11907 Human bre

26	450	100.0	120	2 AAY48606	Aay48606 Human bre
27	450	100.0	182	6 ABP96110	Abp96110 Human mam
28	447	99.3	90	4 AAE07521	Aae07521 Human lip
29	447	99.3	90	6 ABP96098	Abp96098 Human lip
30	445	98.9	90	4 AAE07525	Aae07525 Human lip
31	445	98.9	90	6 ABP96103	Abp96103 Human lip
32	445	98.9	182	6 ABP96109	Abp96109 Human mam
33	440	97.8	90	4 AAE07524	Aae07524 Human lip
34	440	97.8	90	6 ABP96102	Abp96102 Human lip
35	436	96.9	90	4 AAE07522	Aae07522 Human lip
36	436	96.9	90	6 ABP96099	Abp96099 Human lip
37	435	96.7	90	4 AAE07526	Aae07526 Human lip
38	435	96.7	90	6 ABP96101	Abp96101 Human lip
39	434	96.6	89	5 AAO20554	Aao20554 Protein o
40	433	96.2	90	4 AAE07527	Aae07527 Human lip
41	433	96.2	90	6 ABP96104	Abp96104 Human lip
42	417	92.7	88	4 AAE07523	Aae07523 Human lip
43	417	92.7	88	6 ABP96100	Abp96100 Human lip
44	342	76.0	69	2 AAW54279	Aaw54279 BUI01 ant
45	342	76.0	69	3 AAB07509	Abp07509 Amino aci
46	342	76.0	69	8 ADR46908	Adr46908 Breast ca
47	342	76.0	145	6 ABP96111	Abp96111 Human mam
48	342	76.0	145	6 ABP96112	Abp96112 Human mam
49	277	61.6	90	2 AAW35802	Aaw35802 Human end
50	277	61.6	90	3 AAB03767	Abp03767 Human end
51	277	61.6	90	4 AAB31680	Abp31680 An endome
52	277	61.6	90	5 ABB09633	Abp09633 Human end
53	277	61.6	90	6 ABG73167	Abg73167 Human end
54	267	59.3	53	3 AAY64680	Aay64680 Human 5'
55	267	59.3	53	4 AAG00474	Aag00474 Human sec
56	267	59.3	90	4 AAM23955	Aam23955 Human EST
57	267	59.3	90	8 ADS11987	Adsl1987 Human the
58	267	59.3	90	8 ADS10728	Adsl0728 Human the
59	261	58.0	90	5 AAO20553	Aao20553 Protein o
60	254	56.4	102	4 AAB87666	Abb87666 Bovine ma
61	246	54.7	50	4 ABB40007	Abb40007 Peptide #
62	246	54.7	50	4 AAM33628	Aam33628 Peptide #
63	246	54.7	50	4 AAM73426	Aam73426 Human bra
64	246	54.7	50	4 AAM60754	Aam60754 Human bra
65	246	54.7	50	4 ABG55149	Abg55149 Human liv
66	246	54.7	50	5 ABG43285	Abg43285 Human pep
67	244	54.2	79	2 AAY59999	Aay59999 Human end
68	242	53.8	92	4 ABG26041	Abg26041 Novel hum
69	238	52.9	83	3 AAY66704	Aay66704 Membrane-
70	238	52.9	83	3 AAB24417	Aab24417 Human PRO
71	238	52.9	83	4 AAU12391	Aau12391 Human PRO
72	238	52.9	83	4 AAB65227	Abu65227 Human PRO
73	238	52.9	83	6 ABUS8042	Abu58042 Human PRO
74	238	52.9	83	6 ABUS9120	Abu59120 Novel hum
75	238	52.9	83	6 ABUS2632	Abu82632 Human sec
76	238	52.9	83	6 ABO17835	Abol17835 Novel hum
77	238	52.9	83	6 ABU60551	Abu60551 Human sec
78	238	52.9	83	6 ABU13933	Abu13933 Human PRO
79	238	52.9	83	6 ABU81089	Abu81089 Human PRO
80	238	52.9	83	6 ABU72518	Abu72518 Novel hum
81	238	52.9	83	6 ABU66789	Abu66789 Human PRO
82	238	52.9	83	6 ABUS9870	Abu59870 Novel sec
83	238	52.9	83	6 ABUS9267	Abu59267 Human sec
84	238	52.9	83	6 ABO25964	Abol25964 Human PRO
85	238	52.9	83	6 ABO25060	Abol25060 Human sec
86	238	52.9	83	6 ABUS8973	Abu58973 Human sec
87	238	52.9	83	6 ABUS92351	Abu92351 Novel hum
88	238	52.9	83	6 ABUS9416	Abu59416 Novel hum
89	238	52.9	83	6 ABU67065	Abu67065 Human sec
90	238	52.9	83	6 ABU92182	Abu92182 Novel hum
91	238	52.9	83	6 ABU10888	Abu10888 Human PRO
92	238	52.9	83	6 ABU1640	Abu1640 Novel hum
93	238	52.9	83	6 ABUS8579	Abu8579 Human sec
94	238	52.9	83	6 ABO34093	Abol34093 Human PRO
95	238	52.9	83	6 ADA45959	Ada45959 Novel hum
96	238	52.9	83	6 ADA76390	Ada76390 Human PRO
97	238	52.9	83	6 ADA19040	Ada19040 Human PRO
98	238	52.9	83	6 ADA61663	Ada61663 Homo sapi

99 238 52.9 83 6 ADB19448 Adb19448 Novel hum  
100 238 52.9 83 6 ADB27989 Adb27989 Human PRO

## ALIGNMENTS

RESULT 1  
ID AAW35803  
XX AAW35803 standard; protein; 90 AA.  
AC AAW35803;

XX 27-MAR-1998 (first entry)

XX Human endometrial specific steroid-binding factor II.

DE Endometrial specific steroid-binding factor II; ESF II; human;  
KW Clara cell secretory protein; endometrium; phospholipase A2 inhibitor;  
KW polychlorinated biphenyl; antiagregant; inflammation; asthma; rhinitis;  
KW cystic fibrosis; airway disease; neoplasia; atopy; therapy; diagnosis.

OS Homo sapiens.

XX Key Location/Qualifiers  
FH Peptide 1..21  
FT /label= Sig\_peptide  
FT Protein 22..90  
FT /label= Mat\_protein

XX WO9734997-A1.

PN 25-SEP-1997.

XX 21-MAR-1996; 96WO-US003857.

XX 21-MAR-1996; 96WO-US003857.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Yu G, Gentz RL;

XX WPI; 1997-480206/44.

DR N-PSDB; AAT94831.

XX Human endometrial specific steroid-binding factor I, II and III - used to  
PT treat inflammation, asthma, rhinitis, cystic fibrosis, airway disease,  
PT neoplasia, atopy etc.

PS Claim 16; Page 63-64; 92pp; English.

XX This sequence comprises human endometrial specific steroid binding factor  
CC II (ESF II), a protein that inhibits phospholipase A2 activity, binds to  
CC polychlorinated biphenyl compounds, reduces foreign protein antigenicity,  
CC inhibits monocyte and neutrophil chemotaxis and phagocytosis, inhibits  
CC platelet aggregation, regulates eicosanoid levels in the human uterus and  
CC controls the growth of endometrial cells. The amino acid sequence was  
CC deduced from a cDNA clone (see AAT94831) derived from cycloheximide-  
CC treated CEM cells. ESF I (see AAW35802) and ESF III (see AAW35804) are  
CC also claimed. Human ESF II has about 49% identity with rat prostatic  
CC steroid-binding protein. Recombinant ESF I, II and III can be expressed  
CC in host cells for use in claimed methods (a) for treating a patient in  
CC need of ESF I, II or III (including expression of the polypeptide in  
CC vivo) and (b) for identifying compounds which bind to and inhibit  
CC activation of the ESF polypeptide. hESF I, II and III may be used to  
CC treat inflammation, asthma, rhinitis, cystic fibrosis, airway disease,  
CC neoplasia and atopy

XX Sequence 90 AA;

Query Match 100.0%; Score 450; DB 2; Length 90;

Best Local Similarity 100.0%; Pred. No. 7e-50;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLSVCLLIIVTLALCCYQANAEFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAKL 60  
|||  
Db 1 MKLSVCLLIIVTLALCCYQANAEFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAKL 60  
|||  
QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
|||  
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
|||

## RESULT 2

AAW54271

ID AAW54271 standard; protein; 90 AA.

XX AAW54271;

XX 25-MAR-2003 (revised)

DT 29-JUL-1998 (first entry)

XX BU101 antigenic peptide epitope 1.

XX BU101; breast cancer; diagnosis; prevention; treatment; gene therapy;  
KW immunisation; drug screening; epitope.

OS Homo sapiens.

XX WO9807857-A1.

XX 26-FEB-1998.

XX 19-AUG-1997; 97WO-US014665.

XX 19-AUG-1996; 96US-00697105.

PR 15-AUG-1997; 97US-00912276.

XX (ABBO ) ABBOTT LAB.

PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;

PI Granados EN, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp L;

PI Russell JC, Stroupe SD;

XX WPI; 1998-169161/15.

DR N-PSDB; AAV26461.

XX New BU101 protein over-expressed in breast cancer - useful for, e.g.  
PT diagnosis, treatment and prevention of breast cancer.

PS Claim 17; Page 90; 105pp; English.

XX This represents a BU101 polypeptide sequence. BU101 is a member of the  
CC uteroglobin family of proteins and is over-expressed in breast cancer.  
CC Cells transformed with a recombinant expressing system comprising a  
CC sequence derived from the BU101 open reading frame and with at least 50  
CC percent identity to the sequences shown in AAV26458 to AAV26461 are used  
CC to produce BU101 polypeptides containing at least 1 epitope. These are  
CC used to detect BU101-specific antibodies which are used correspondingly  
CC to detect BU101 antigens. The BU101 polynucleotide sequences can be used  
CC in a method for detecting the presence of a target BU101 polynucleotide.  
CC The various assays are used for diagnosis, prognosis, staging,  
CC monitoring, treating and preventing diseases of the breast (especially  
CC cancer and its metastases), and also for determining susceptibility. The  
CC BU101 polypeptides are also useful in drug screening, e.g. to identify  
CC antagonists of BU101, potentially useful therapeutically and as targets  
CC for therapy. The antibodies are also useful for targeted drug delivery  
CC and therapeutically to neutralise BU101 polypeptides. Fragments of the  
CC BU101 nucleic acid are useful as probes and primers, e.g. for detection  
CC of altered gene expression or in fluorescent in situ hybridisation, also  
CC in gene therapy to generate antisense or ribozyme molecules or for  
CC genetic immunisation. (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 90 AA;

Query Match 100.0%; Score 450; DB 2; Length 90;



Best Local Similarity 100.0%; Pred. No. 7e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLSVCLLLVTLALCCYQANAEFCFALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCFALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60

Qy 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

RESULT 3  
AAW59776  
ID AAW59776 standard; protein; 90 AA.  
XX  
AC AAW59776;  
XX  
DT 12-OCT-1998 (first entry)  
XX  
DE Amino acid sequence of the human steroid binding protein C1.  
XX  
KW Human steroid-binding protein C1; hSBP1; hSBP2; breast cancer; probe;  
KW gene therapy vector; ribozyme; probe; hybridisation; amplification;  
KW antibody; immunoassay.  
XX  
OS Homo sapiens.  
XX  
PN W09821331-A1.  
XX  
PD 22-MAY-1998.  
XX  
PF 07-NOV-1997; 97WO-US020674.  
XX  
PR 12-NOV-1996; 96US-00747547.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Akerblom IE, Hillman JL, Murry LE, Goli SK, Hawkins PR;  
XX  
DR WPI; 1998-297935/26.  
DR N-PSDB; AAV41579.  
XX  
XX New human steroid binding proteins C1 and C2 - useful for, e.g.  
diagnosis, monitoring and treating breast cancer, and for drug screening.  
PS Claim 1; Fig 1; 70pp; English.  
XX  
CC This is the amino acid sequence of the human steroid-binding protein C1  
(hSBP1) used in the method of the invention for the diagnosis, monitoring  
CC and treatment of breast cancer. hSBP1 and hSBP2 are useful as markers for  
CC breast cancer, i.e. measuring levels of hSBP1 and hSBP2 used for  
CC diagnosis or monitoring the disease, to identify subjects at risk and to  
CC discriminate between different forms of cancer for selection of  
CC appropriate therapies. They may also be used for drug screening. Nucleic  
CC acids encoding hSBP1 and hSBP2 can be used in gene therapy vectors to  
CC overexpress the steroid-binding proteins, preventing binding of steroids,  
CC or antisense sequences, ribozymes. Their nucleic acids can also be used,  
CC for the diagnosis and monitoring (by quantifying expression of hSBP), as  
CC source of probes for hybridisation and amplification of genomic or  
CC related sequences for studying regulation of gene function and for  
CC mapping the genomic sequence. Antibodies are used as diagnostic reagents  
CC in standard immunoassays for hSBP  
XX  
SQ Sequence 90 AA;  
Query Match 100.0%; Score 450; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 7e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLSVCLLLVTLALCCYQANAEFCFALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCFALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60

Qy 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

RESULT 5  
AAW84875

Qy 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

RESULT 4  
AAW89613  
ID AAW89613 standard; protein; 90 AA.  
XX  
AC AAW89613;  
XX  
DT 25-MAR-1999 (first entry)  
XX  
DE Endometrial steroid binding protein II.  
XX  
KW Endometrial steroid binding protein II; ESBPII; cancer; detection;  
KW endometriosis; endometrial fibroid; mammary cancer.  
XX  
OS Homo sapiens.  
XX  
PN W09856248-A1.  
XX  
PD 17-DEC-1998.  
XX  
PF 09-JUN-1998; 98WO-US012053.  
XX  
PR 09-JUN-1997; 97US-0049015P.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PI Schmidt CJ, Wang X;  
XX  
DR WPI; 1999-080843/07.  
DR N-PSDB; AAX00069.  
XX  
PT Treatment of endometrial cancer, mammary cancer, endometriosis or  
endometrial fibroids - comprises administering endometrial steroid  
binding protein II antagonist.  
XX  
PS Disclosure; Page 13; 19pp; English.  
XX  
CC A method has been developed for the treatment of endometrial cancer,  
CC mammary cancer, endometriosis or endometrial fibroids. The method  
CC comprises administering endometrial steroid binding protein II (ESBPII)  
CC antagonist. Also described in the present invention are: (1) a method for  
CC diagnosing the above mentioned diseases comprising analysing the  
CC abnormally high level of ESBPII polypeptide in cells, tissues and bodily  
CC fluids; and (2) a diagnostic method for the diseases described above  
CC comprising analysing the abnormally high or low transcription level of  
CC ESBPII in cells, tissues and bodily fluids. The methods can be used to  
CC diagnose, treat, and monitor the progression, remission or recurrence of  
CC abnormal cell growth, such as cancers, especially endometrial and mammary  
CC cancer, and endometriosis and endometrial fibroids. The present sequence  
CC represents ESBPII, from the present invention  
XX  
SQ Sequence 90 AA;

Query Match 100.0%; Score 450; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 7e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLSVCLLLVTLALCCYQANAEFCFALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCFALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60

Qy 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

ID AAY84875 standard; protein; 90 AA.  
 AC AAY84875;  
 XX  
 DT 08-AUG-2000 (first entry)  
 XX  
 XX  
 DE A human endometrial specific steroid-binding protein II.  
 XX  
 KW Human; endometrial specific steroid-binding protein II; ESBPII;  
 KW breast tumour; prostate cancer; gynaecological cancer; cancer;  
 KW endometrial cancer; ovarian cancer; uterine cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200020043-A1.  
 PN  
 XX  
 PD 13-APR-2000.  
 XX  
 XX  
 PF 05-OCT-1999; 99WO-US023252.  
 XX  
 XX 05-OCT-1998; 98US-0103093P.  
 PR  
 XX (DIAD-) DIADEXUS LLC.  
 PA  
 XX Macina RA;  
 PI  
 XX WPI; 2000-303648/26.  
 DR N-PSDB; AAA14953.  
 DR  
 XX Diagnosing, staging, monitoring, imaging and treating prostate and  
 PT gynecological cancers by measuring levels of endometrial specific steroid  
 PT -binding protein (ESBPII) expression.  
 PT  
 XX Claim 6; Page 31-32; 35pp; English.  
 PS  
 XX The present sequence represents a human endometrial specific steroid-  
 CC binding protein (ESBP) II. The ESBPII protein is overexpressed in breast  
 CC tumours. The specification describes a method for diagnosing prostate or  
 CC a gynaecological cancer. The method comprises measuring levels of ESBPII  
 CC in cells, tissues or body fluids of a patient, and comparing this to  
 CC levels from a normal control, where a variance in levels indicates  
 CC cancer. The method is used to diagnose, stage, monitor, image or treat  
 CC prostate or gynaecological cancer. The gynaecological cancers include  
 CC breast, endometrial, ovarian and uterine cancer  
 XX  
 SQ Sequence 90 AA;  
 Query Match 100.0%; Score 450; DB 3; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60  
 DB 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60  
 QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 DB 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 RESULT 6  
 AAB03768  
 ID AAB03768 standard; protein; 90 AA.  
 XX  
 AC AAB03768;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 XX Human endometrial specific steroid-binding factor II protein sequence.  
 DE  
 DE Endometrial specific steroid-binding factor; human; hESF; inflammation;  
 KW asthma; rhinitis; cystic fibrosis; air way disease; neoplasia; acopy;  
 KW eicosanoid level regulator; chemotaxis inhibitor; endometrial cancer; ss.  
 KW

XX Homo sapiens.  
 OS  
 XX US6066724-A.  
 PN  
 XX 23-MAY-2000.  
 PD  
 XX 21-MAR-1997; 97US-00821451.  
 PF  
 XX 21-MAR-1996; 96US-0014724P.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Yu G, Gentz R, Ni J;  
 PI WPI: 2000-375600/32.  
 XX N-PSDB; AAA59729.  
 DR  
 DR Novel gene encoding human endometrial specific steroid-binding factor I,  
 PT II and III which is useful for treating asthma, rhinitis, cystic  
 PT fibrosis, airway disease and neoplasia.  
 XX  
 PS Claim 1; Fig 2; 36pp; English.  
 XX  
 CC This invention relates to nucleic acid molecules encoding portions of the  
 CC human endometrial specific steroid-binding factors I, II, and III. Also  
 CC included in the invention are hESF I, II, and III polypeptide sequences.  
 CC The nucleotide sequence exhibit antiasthmatic, antiinflammatory,  
 CC antiallergic, and cytostatic properties. The polynucleotides are used in  
 CC gene therapy to express hESF I, II and III polypeptides in vivo to treat  
 CC disease, neoplasia and atopy. The polynucleotides are also used to  
 CC inhibit phospholipase A2 activity, bind polychlorinated biphenyls, reduce  
 CC foreign protein antigenicity, inhibit monocyte and neutrophil chemotaxis  
 CC and phagocytosis, inhibit platelet aggregation, regulate eicosanoid  
 CC levels in the human uterus and control the growth of endometrial cells.  
 CC The polynucleotides are also useful for detecting complementary  
 CC polynucleotides as a diagnostic reagent. The hESF I, II and III  
 CC polynucleotides are used to detect complementary polynucleotides such as  
 CC a diagnostic reagent. Detection of a mutated form of hESF I, II and III  
 CC associated with a dysfunction will provide a diagnostic tool that can  
 CC define diagnosis of a disease or susceptibility to a disease which  
 CC results from under-expression, over-expression or altered expression of  
 CC hESF I, II and III e.g. a susceptibility to inherited asthma and  
 CC endometrial cancer. They are also useful for chromosome identification.  
 CC The present sequence represents a hESF II protein sequence identified in  
 CC the invention  
 XX  
 SQ Sequence 90 AA;  
 Query Match 100.0%; Score 450; DB 3; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60  
 DB 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60  
 QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 DB 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 RESULT 7  
 AAB13787  
 ID AAB13787 standard; protein; 90 AA.  
 XX  
 AC AAB13787;  
 XX  
 DT 20-JUN-2001 (first entry)  
 XX  
 XX Human BU101.  
 DE

KW Human; breast cancer; breast disease detection; mammaglobin; uteroglobin;  
 KW BU101; endometrial; cytostatic.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 53  
 FT /label= Leu  
 FT /note= "Encoded by CTG in polymorphic variant"  
 XX WO200035950-A2.  
 PN  
 XX 22-JUN-2000.  
 XX  
 XX 20-DEC-1999; 99WO-US030489.  
 XX  
 XX 18-DEC-1998; 98US-00215818.  
 XX  
 XX (ABBO ) ABBOTT LAB.  
 XX  
 XX Colpitts TL, Russell JE;  
 PI WPI; 2000-442366/38.  
 DR N-PSDB; AAA64846.  
 DR  
 XX  
 XX Multimeric polypeptide antigen and antibody specific to the antigen are  
 PT useful for diagnosing, detecting and treating breast cancer.  
 PT  
 XX  
 XX Claim 1; Page 124; 124pp; English.  
 XX  
 XX BU101 is a member of the uteroglobin protein family. The present sequence  
 CC is the protein sequence for human BU101. The present invention relates to  
 CC a multimeric polypeptide antigen, which comprises of the present sequence  
 CC and mammaglobin polypeptide (AA013786). Mammaglobin is another  
 CC uteroglobin protein. The presence of multimeric polypeptide antigen in a  
 CC test sample can be used as the basis for a test to diagnose breast  
 CC disease e.g. breast cancer, in a patient. The detection can be carried  
 CC out using antibodies specific for the multimeric polypeptide antigen. The  
 CC present sequence can either have a Pro or Leu residue at position 53,  
 CC since the coding sequence has a single nucleotide T/C polymorphism at  
 CC nucleotide position 254  
 XX  
 XX Sequence 90 AA;  
 SQ  
 Query Match 100.0%; Score 450; DB 3; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
 Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
 QY 61 GVKRCTDQMSLOKRSLIAEVLVKLKCSV 90  
 Db 61 GVKRCTDQMSLOKRSLIAEVLVKLKCSV 90  
 RESULT 8  
 AAB07501  
 ID AAB07501 standard; protein; 90 AA.  
 XX  
 XX AAB07501;  
 AC  
 XX  
 DT 20-OCT-2000 (first entry)  
 XX  
 XX Amino acid sequence of a human BU101 polypeptide.  
 DE  
 XX Human; BU101; breast disease.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200041516-A2.  
 PN  
 XX

PD 20-JUL-2000.  
 XX 19-JAN-2000; 2000WO-US001309.  
 PF  
 XX 19-JAN-1999; 99US-00233693.  
 PR  
 XX (ABBO ) ABBOTT LAB.  
 XX  
 XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;  
 PI Granados EN, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp L;  
 PI Russell JC, Scheffel CP, Stroupe SD;  
 XX WPI; 2000-475906/41.  
 DR N-PSDB; AAA58880.  
 DR  
 XX Detecting presence of target BU101 polynucleotide in sample useful for  
 PT detection of breast cancer, comprises contacting sample with BU101-  
 PT specific polynucleotide and determining binding.  
 XX  
 XX Claim 23; Page 125; 127pp; English.  
 XX  
 XX The present sequence represents a BU101 polypeptide. The BU101 gene is  
 CC transcribed from breast tissue. The specification describes a method for  
 CC detecting the presence of a target BU101 polynucleotide in a test sample.  
 CC The method comprises contacting the sample with at least one BU101-  
 CC specific polynucleotide (AAA58875-80), and detecting bound  
 CC polynucleotides. The method and BU101 polynucleotides are useful for  
 CC detecting the presence of BU101 polynucleotides. The methods may be used  
 CC for the diagnosis of breast disease, indicated by the formation of  
 CC complexes  
 XX  
 XX Sequence 90 AA;  
 SQ  
 Query Match 100.0%; Score 450; DB 3; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
 Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
 QY 61 GVKRCTDQMSLOKRSLIAEVLVKLKCSV 90  
 Db 61 GVKRCTDQMSLOKRSLIAEVLVKLKCSV 90  
 RESULT 9  
 AAG65989  
 ID AAG65989 standard; protein; 90 AA.  
 XX  
 XX AAG65989;  
 AC  
 XX  
 DT 11-FEB-2002 (first entry)  
 XX  
 XX Lipophilin B polypeptide.  
 DE  
 XX Genetic subtraction; DNA microarray analysis; polymerase chain reaction;  
 KW cancer; B726P; Lipophilin B; mammaglobin.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200175171-A2.  
 PN  
 XX  
 XX 11-OCT-2001.  
 PD  
 XX 02-APR-2001; 2001WO-US010631.  
 PF  
 XX 03-APR-2000; 2000US-0194241P.  
 PR  
 XX 20-JUL-2000; 2000US-0219862P.  
 PR  
 XX 27-JUL-2000; 2000US-0221300P.  
 PR  
 XX 18-DEC-2000; 2000US-0256592P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA

XX Houghton RL, Dillon DC, Molesh DA, Xu J, Zehentner B, Persing DH;  
 XX WPI; 2001-626449/72.  
 DR N-PSDB; AAI67269.  
 XX Identifying tissue (tumor)-specific polynucleotides overexpressed in  
 PT tissue of interest as compared to control tissue, for detecting cancer  
 PT cells in patient, comprises DNA microarray analysis or quantitative  
 PT polymerase chain reaction.  
 XX Example; Page 127; 127pp; English.  
 XX The invention relates to identifying tissue-specific polynucleotides (P)  
 CC that involves performing a genetic subtraction to identify pool of (P)  
 CC from tissue of interest (TI), performing DNA microarray analysis to  
 CC identify first subset of polynucleotides (SPI) at least 2-fold over  
 CC expressed in TI, and performing quantitative polymerase chain reaction  
 CC (PCR) analysis on SPI to identify second subset of (P). The method is  
 CC useful for determining the presence or absence of a cancer cell in a  
 CC patient, monitoring the progression of cancer in a patient using a  
 CC biological sample such as blood, serum, lymph nodes, bone marrow, sputum,  
 CC urine or a tumour biopsy sample. The methods are useful for determining  
 CC the presence or absence of or monitoring progression of prostate, breast,  
 CC colon, ovarian, lung, head and neck, lymphoma, leukemia, melanoma, liver,  
 CC gastric, kidney, bladder, pancreatic or endometrial cancer. The present  
 CC sequence represents the lipophilin B polypeptide  
 XX Sequence 90 AA;  
 SQ Query Match 100.0%; Score 450; DB 4; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQNAEFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAKL 60  
 Db 1 MKLSVCLLLVTLALCCYQNAEFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAKL 60  
 QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 RESULT 10  
 AAB31681  
 ID AAB31681 standard; protein; 90 AA.  
 XX AAB31681;  
 AC AAB31681;  
 XX 30-APR-2001 (first entry)  
 DT An endometrial specific steroid binding factor II.  
 DE Human, endometrial specific steroid binding factor; hESF; hESFI; hESFII;  
 XX hESFIII; inflammation; asthma; rhinitis; cystic fibrosis; airway disease;  
 KW neoplasia; atopy; phospholipase A2; polychlorinated biphenyl; chemotaxis;  
 KW phagocytosis; platelet aggregation; eicosanoid; endometrial cell.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FH Peptide 1..21  
 FT /note= "signal peptide"  
 XX US6174992-B1.  
 PN 16-JAN-2001.  
 PD 08-MAR-1999; 99US-00263810.  
 XX 21-MAR-1996; 96US-0014724P.  
 PR 21-MAR-1997; 97US-00821451.  
 XX

(HUMA-) HUMAN GENOME SCI INC.  
 XX Ni J, Yu G, Gentz R;  
 XX WPI; 2001-158477/16.  
 DR N-PSDB; AAF25213.  
 XX New human endometrial specific steroid binding factors, useful for  
 PT treating and preventing inflammation, asthma, rhinitis, cystic fibrosis,  
 PT airway disease, neoplasia and atopy.  
 XX Claim 1; Fig 2; 36pp; English.  
 XX The present sequence represents a human endometrial specific steroid  
 CC binding factor (hESF). The specification describes hESFI, hESFII, and  
 CC hESFIII. hESFI, II and III polypeptides, and polynucleotides encoding  
 CC them are useful for treating and preventing inflammation, asthma,  
 CC rhinitis, cystic fibrosis, airway disease, neoplasia and atopy,  
 CC inhibiting phospholipase A2 activity, binding polychlorinated biphenyls,  
 CC reducing foreign protein antigenicity, inhibiting monocyte and neutrophil  
 CC chemotaxis and phagocytosis, inhibiting platelet aggregation, regulating  
 CC eicosanoid levels in the human uterus, and for controlling the growth of  
 CC endometrial cells. hESF polypeptides and nucleotides are also useful for  
 CC research, biological, clinical or therapeutic purposes  
 XX Sequence 90 AA;  
 SQ Query Match 100.0%; Score 450; DB 4; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQNAEFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAKL 60  
 Db 1 MKLSVCLLLVTLALCCYQNAEFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAKL 60  
 QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
 RESULT 11  
 AAE07518  
 ID AAE07518 standard; protein; 90 AA.  
 XX AAE07518;  
 AC AAE07518;  
 XX 06-NOV-2001 (first entry)  
 DT Human lipophilin B protein.  
 DE Human; lipophilin B; cytostatic; vaccine; gene therapy; uteroglobin;  
 KW cancer; breast; ovary; prostate.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200158947-A1.  
 PN 16-AUG-2001.  
 PD 08-FEB-2001; 2001WO-US004439.  
 PF 11-FEB-2000; 2000US-0183495P.  
 PR 28-JUN-2000; 2000US-0215735P.  
 XX (CORI-) CORIXA CORP.  
 PA Carter D, Vedvick TS, Vallieve-Douglass J, Houghton RL, Dillon DC;  
 XX WPI; 2001-497069/54.  
 DR N-PSDB; AAD13756.  
 XX Novel isolated complex two lipophilin-like polypeptides linked by at  
 PT least one disulfide bond, used to treat or prevent breast, ovarian or

PT prostate cancer.  
XX Example 5; Page 72; 91pp; English.  
PS  
XX The invention relates to a complex comprising a lipophilin-like polypeptide linked by at least one disulphide bond to a second lipophilin-like polypeptide. Lipophilin-like proteins are members of uteroglobin superfamily. Lipophilin-like proteins are useful in the preparation of vaccines. The complex containing lipophilin-like proteins are useful for treating or preventing breast, ovarian or prostate cancer. The complex is also used for determining the presence or absence of cancer in a patient, or monitor the progress of cancer in a patient. Lipophilin DNA is also useful in gene therapy. The present sequence is human lipophilin B protein  
XX  
SQ Sequence 90 AA;  
Query Match 100.0%; Score 450; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 7e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
Qy 61 GVKRCTDQMSLQKRSLSIAEVLVKLKCSV 90  
Db 61 GVKRCTDQMSLQKRSLSIAEVLVKLKCSV 90  
RESULT 12  
ABG96363  
ID ABG96363 standard; protein; 90 AA.  
XX  
AC ABG96363;  
DT 11-DEC-2002 (first entry)  
XX  
DE Human ovarian cancer marker M559.  
XX  
KW Human; ovarian cancer; marker; cancer; familial history; brain disorder; central nervous system disorder; bacterial meningitis; viral meningitis; Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus; brain herniation; inflammation; encephalitis; testicular disorder; nontuberculous granulomatous orchitis; connective tissue disorder; heart disorder; ischaemic heart disease; atherosclerosis; neoplasm; histological type; carcinogenic; ovarian cancer marker.  
XX  
OS Homo sapiens.  
XX  
XX WO200271928-A2.  
XX  
XX 19-SEP-2002.  
XX  
XX 14-MAR-2002; 2002WO-US007826.  
XX  
XX 14-MAR-2001; 2001US-0276025P.  
XX  
XX 14-MAR-2001; 2001US-0276026P.  
XX  
XX 10-AUG-2001; 2001US-0311732P.  
XX  
XX 19-SEP-2001; 2001US-0323580P.  
XX  
XX 26-SEP-2001; 2001US-0324967P.  
XX  
XX 26-SEP-2001; 2001US-0325102P.  
XX  
XX 26-SEP-2001; 2001US-0325149P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG; Meyers RE, Morrissey MF, Olandt PJ, Sen A, Vieby PO, Mills GB; Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;  
XX  
XX WPI; 2002-723277/78.  
XX  
XX N-PSDB; ABS76459.  
XX  
PT Assessing whether a patient is afflicted with ovarian cancer, useful in assessing the stage or progression of the disease, comprises comparing the expression level of a cancer marker in a sample from a patient and from a non cancer patient.  
PT  
XX Disclosure; Page 299-300; 481pp; English.  
XX  
XX The present invention relates to a new method for assessing whether a patient is afflicted with ovarian cancer. The method involves comparing the expression level of a marker in a patient sample and the normal level of expression of the marker in a control non-ovarian cancer sample, where the marker is selected from 363 cancer markers described in the specification. The method of the invention is useful in diagnosing or characterising cancer, in detecting the presence of cancer as early as possible, and the recurrence of ovarian cancer. The method may also be of particular use with patients having an enhanced risk of developing ovarian cancer (e.g. patients having a familial history of ovarian cancer). The cancer markers may be used in the management and treatment of e.g. brain and central nervous system disorders (e.g. bacterial and viral meningitis, Alzheimer's disease or Parkinson's disease), brain disorders (e.g. cerebral oedema, hydrocephalus or brain herniations), inflammations (e.g. bacterial or viral meningitis or encephalitis), testicular disorders (e.g. nontuberculous granulomatous orchitis), connective tissue disorders, or heart disorders (e.g. ischaemic heart disease or atherosclerosis). The compositions and methods may also be used in assessing the histological type of neoplasm associated with ovarian cancer, monitoring the progression of ovarian cancer, determining whether ovarian cancer has metastasized or is likely to metastasize, selecting a composition for inhibiting ovarian cancer, assessing the ovarian carcinogenic potential of a compound, or inhibiting ovarian cancer or at risk of developing ovarian cancer. The present amino acid sequence represents one of the ovarian cancer markers described in the invention  
XX  
SQ Sequence 90 AA;  
Query Match 100.0%; Score 450; DB 5; Length 90;  
Best Local Similarity 100.0%; Pred. No. 7e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAL 60  
Qy 61 GVKRCTDQMSLQKRSLSIAEVLVKLKCSV 90  
Db 61 GVKRCTDQMSLQKRSLSIAEVLVKLKCSV 90  
RESULT 13  
ABG09634  
ID ABG09634 standard; protein; 90 AA.  
XX  
AC ABG09634;  
XX  
XX 29-MAY-2002 (first entry)  
XX  
XX Human endometrial specific steroid-binding factor (hESF) II.  
XX  
XX Human; endometrial specific steroid-binding factor; ESF; prostatic steroid-binding protein; hESF I; hESF II; hESF III; asthma.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Key 1..21  
XX Peptide /note= "signal peptide"  
XX Protein 22..90  
XX /note= "mature protein"  
XX  
XX US6338948-B1.  
XX  
XX 15-JAN-2002.  
XX

```
XX PF 30-MAY-2000; 2000US-00583169.
XX XX
XX PR 21-MAR-1996; 96US-0014724P.
XX PR 21-MAR-1997; 97US-00821451.
XX PR 08-MAR-1999; 99US-00263810.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Ni J, Yu G, Gentz R;
XX XX
XX DR WPI; 2002-215019/27.
XX DR N-PSDB; ABL41782.
XX XX
XX PT New antibody specific for human endometrial specific steroid-binding
XX PT factor (hESF) III, useful for detecting hESF III protein in biological
XX PT sample and to isolate or identify clones expressing the protein.
XX XX
XX PS Disclosure; Fig 1; 36pp; English.
XX XX
XX CC The present sequence represents a endometrial specific steroid-binding
XX CC factor (hESF) II. The full length protein has a molecular weight of 9.9
XX CC kDa. The protein has homology to rat prostatic steroid-binding protein
XX CC C2. Antibodies which bind hESF proteins, such as hESF I, hESF II, and
XX CC hESF III are useful for isolating or to identify clones expressing the
XX CC polypeptides or to purify the polypeptides by affinity chromatography.
XX CC Agonists and antagonists of hESF proteins are useful for treating and/or
XX CC preventing susceptibility to asthma
XX XX
XX SQ Sequence 90 AA;
XX
XX Query Match 100.0%; Score 450; DB 5; Length 90;
XX Best Local Similarity 100.0%; Pred. No. 7e-50;
XX Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFIISBPLFKLSIAKFDAPPEAVAAGL 60
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFIISBPLFKLSIAKFDAPPEAVAAGL 60
QY 61 GVKRCTDQMSLQKRSIAEVLVKILKCSV 90
Db 61 GVKRCTDQMSLQKRSIAEVLVKILKCSV 90
RESULT 14
ABJ05577
ID ABJ05577 standard; protein; 90 AA.
XX AC ABJ05577;
XX XX
XX DT 14-NOV-2002 (first entry)
XX DE Breast cancer-associated protein 42.
XX XX
XX KW Breast cancer; breast cancer-associated gene sequence; drug development;
XX KW pharmacogenetics; biosensor development.
XX XX
XX OS Unidentified.
XX XX
XX PN WO200259377-A2.
XX XX
XX PD 01-AUG-2002.
XX XX
XX PF 24-JAN-2002; 2002WO-US002242.
XX XX
XX PR 24-JAN-2001; 2001US-0263965P.
XX PR 02-FEB-2001; 2001US-0265928P.
XX PR 09-APR-2001; 2001US-00829472.
XX PR 09-APR-2001; 2001US-0282698P.
XX PR 04-MAY-2001; 2001US-0288590P.
XX PR 29-MAY-2001; 2001US-0294443P.
XX XX
XX PA (BOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Mack DH, Gish KC, Afar D;
XX WPI; 2002-583738/52.
XX N-PSDB; ABT07734.
XX
XX Detecting a breast cancer-associated transcript in a patient's cell,
XX useful for diagnosing breast cancer, comprises contacting a biological
XX sample with a polynucleotide that selectively hybridizes with breast
XX cancer nucleic acids.
XX
XX Disclosure; Page 385; 414pp; English.
XX
XX The invention comprises a method of detecting a breast cancer-associated
XX transcript in a cell from a patient. The method of the invention involves
XX contacting a biological sample from the patient with a nucleotide that
XX hybridizes to one of the 69 breast cancer-associated gene sequences shown
XX in the specification. The method of the invention is useful in the
XX diagnosis or prognosis of breast cancer, and for detecting genes that are
XX up or down-regulated in breast cancer cells. Genes identified by the
XX method of the invention can be used in diagnostic purposes and also as
XX targets for screening for therapeutic compounds that modulate breast
XX cancer (e.g. hormones or antibodies). Identification of genes that are
XX over or under expressed in breast cancer can additionally provide high-
XX resolution, high-sensitivity datasets which can be used in the areas of
XX diagnostic, therapeutic, drug development, pharmacogenetics, protein
XX structure and biosensor development. Amino acid sequences ABJ05536 -
XX ABJ05604 represent the proteins encoded by the 69 breast cancer-
XX associated genes of the invention
XX
XX SQ Sequence 90 AA;
XX
XX Query Match 100.0%; Score 450; DB 5; Length 90;
XX Best Local Similarity 100.0%; Pred. No. 7e-50;
XX Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFIISBPLFKLSIAKFDAPPEAVAAGL 60
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFIISBPLFKLSIAKFDAPPEAVAAGL 60
QY 61 GVKRCTDQMSLQKRSIAEVLVKILKCSV 90
Db 61 GVKRCTDQMSLQKRSIAEVLVKILKCSV 90
RESULT 15
ABJ19810
ID ABJ19810 standard; protein; 90 AA.
XX AC ABJ19810;
XX XX
XX DT 10-APR-2003 (first entry)
XX DE Androgen-independent prostate cancer-related protein - SEQ ID No 28.
XX XX
XX KW Androgen-independent cancer; androgen ablation therapy; prostate cancer;
XX KW androgen-dependent prostate cancer; prostate cancer.
XX XX
XX OS Unidentified.
XX XX
XX PN WO200298358-A2.
XX XX
XX PD 12-DEC-2002.
XX XX
XX PF 04-JUN-2002; 2002WO-US017594.
XX XX
XX PR 04-JUN-2001; 2001US-0295917P.
XX PR 13-NOV-2001; 2001US-0350666P.
XX PR 29-MAR-2002; 2002US-0368689P.
XX PR 12-APR-2002; 2002US-0372246P.
XX PR 31-MAY-2002; 2002US-00160233.
XX XX
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.
```

XX AFAR DEH, Agus D, Mack DH;  
 XX WPI; 2003-148602/14.  
 XX Detecting an androgen-independent prostate cancer cell in a sample or  
 PT diagnosing androgen-dependent prostate cancer, by determining the  
 PT presence or absence of genes whose expressions are up- or down-regulated.  
 XX Claim 1; Page 193; 210pp; English.  
 XX The invention comprises a method for detecting an androgen-independent  
 CC cancer cell in a sample from a patient who has undergone androgen  
 CC ablation therapy. The method involves determining the presence or absence  
 CC of nucleic acids that are either up-regulated or down-regulated in  
 CC prostate cancer. The method is useful for detecting an androgen-  
 CC independent prostate cancer cell in a sample from a patient who has  
 CC undergone androgen ablation therapy. The method is particularly useful  
 CC for diagnosing androgen-dependent prostate cancer, prostate cancer  
 CC undergoing androgen withdrawal, or androgen-independent prostate cancer.  
 CC The present amino acid sequence represents a protein which is encoded by  
 CC a gene that is either up-regulated or down-regulated in prostate cancer  
 XX SQ Sequence 90 AA;  
 Query Match 100.0%; Score 450; DB 6; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQANAFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVNAKL 60  
 DB 1 MKLSVCLLLVTLALCCYQANAFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVNAKL 60  
 QY 61 GVKRCTDOMSLQKRSLIAEVLVKILKCSV 90  
 DB 61 GVKRCTDOMSLQKRSLIAEVLVKILKCSV 90  
 RESULT 16  
 ABP96089  
 ID ABP96089 standard; protein; 90 AA.  
 XX AC ABP96089;  
 XX DT 08-MAY-2003 (first entry)  
 XX DE Human lipophilin B amino acid sequence SEQ ID NO:2.  
 XX Human; lipophilin; mammaglobin; breast tumour antigen; fusion protein;  
 KW cancer; cytostatic; vaccine; antibody therapy; tumour.  
 XX OS Homo sapiens.  
 XX WO2003005888-A2.  
 XX 23-JAN-2003.  
 XX 11-JUL-2002; 2002WO-US022325.  
 XX 13-JUL-2001; 2001US-00905673.  
 PR 12-MAR-2002; 2002US-00096319.  
 XX (CORI-) CORIXA CORP.  
 XX Fanger GR, Durham M, Houghton RL, Dillon DC, Carter D;  
 PI Persing DH;  
 XX WPI; 2003-229428/22.  
 XX Fusion protein useful for preventing, treating breast, ovarian and  
 PT prostate cancer, has lipophilin-like polypeptide linked to another  
 PT lipophilin-like polypeptide or a breast tumor antigen by a peptide bond.

PS Example 5; Page 105-106; 132pp; English.  
 XX The present invention describes a fusion protein (FP1) comprising a first  
 CC lipophilin-like polypeptide linked by a peptide bond to a second  
 CC lipophilin-like polypeptide, or a fusion protein (FP2) comprising a first  
 CC lipophilin-like polypeptide linked by a peptide bond to a breast tumour  
 CC antigen. Also described: (1) a polynucleotide (I) encoding FP1 or FP2,  
 CC (2) an isolated antibody (II) or its antigen-binding fragment that  
 CC specifically binds to FP1 or FP2, and does not detectably bind to  
 CC mammaglobin, lipophilin A, B or C, or B305D; and (3) an isolated antibody  
 CC (III) or its antigen-binding fragment that specifically binds to a  
 CC lipophilin complex. FP1 and FP2 have cytostatic activities and can be  
 CC used in vaccines and in antibody therapy. (II) specific for FP1 or (III)  
 CC can be used for determining the presence or absence of breast, ovarian or  
 CC prostate cancer in a patient. FP1, FP2, (I) and (II) can be used as  
 CC vaccines for inhibiting the development of cancer such as breast, ovarian  
 CC or prostate cancer in a patient or a patient considered at risk for such  
 CC a disease may be treated prophylactically. The fusion protein is useful  
 CC for removing tumour cells from a biological sample, and for stimulating  
 CC and/or expanding T cells specific for a lipophilin complex. The fusion  
 CC proteins are useful as markers for progression of cancer. (II) or (III)  
 CC are also useful for monitoring the progression of cancer in a patient.  
 CC The present sequence represents a human lipophilin B protein which is  
 CC used in an example from the present invention  
 XX SQ Sequence 90 AA;  
 Query Match 100.0%; Score 450; DB 6; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 7e-50;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQANAFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVNAKL 60  
 DB 1 MKLSVCLLLVTLALCCYQANAFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVNAKL 60  
 QY 61 GVKRCTDOMSLQKRSLIAEVLVKILKCSV 90  
 DB 61 GVKRCTDOMSLQKRSLIAEVLVKILKCSV 90  
 RESULT 17  
 ABP96097  
 ID ABP96097 standard; protein; 90 AA.  
 XX AC ABP96097;  
 XX DT 08-MAY-2003 (first entry)  
 XX DE Human lipophilin B protein sequence SEQ ID NO:35.  
 XX Human; lipophilin; mammaglobin; breast tumour antigen; fusion protein;  
 KW cancer; cytostatic; vaccine; antibody therapy; tumour.  
 XX OS Homo sapiens.  
 XX OS Synthetic.  
 XX WO2003005888-A2.  
 XX 23-JAN-2003.  
 XX 11-JUL-2002; 2002WO-US022325.  
 XX 13-JUL-2001; 2001US-00905673.  
 PR 12-MAR-2002; 2002US-00096319.  
 XX (CORI-) CORIXA CORP.  
 XX Fanger GR, Durham M, Houghton RL, Dillon DC, Carter D;  
 PI Persing DH;  
 XX WPI; 2003-229428/22.  
 XX Fusion protein useful for preventing, treating breast, ovarian and

PT prostate cancer, has lipophilin-like polypeptide linked to another  
PT lipophilin-like polypeptide or a breast tumor antigen by a peptide bond.  
XX  
PS Example 5; Page 114-115; 132pp; English.  
XX  
CC The present invention describes a fusion protein (FP1) comprising a first  
CC lipophilin-like polypeptide linked by a peptide bond to a second  
CC lipophilin-like polypeptide, or a fusion protein (FP2) comprising a first  
CC lipophilin-like polypeptide linked by a peptide bond to a breast tumor  
CC antigen. Also described: (1) a polynucleotide (I) encoding FP1 or FP2;  
CC (2) an isolated antibody (II) or its antigen-binding fragment that  
CC specifically binds to FP1 or FP2, and does not detectably bind to  
CC mammaglobin, lipophilin A, B or C, or B305D; and (3) an isolated antibody  
CC (III) or its antigen-binding fragment that specifically binds to a  
CC lipophilin complex. FP1 and FP2 have cytostatic activities and can be  
CC used in vaccines and in antibody therapy. (II) specific for FP1 or (III)  
CC can be used for determining the presence or absence of breast, ovarian or  
CC prostate cancer in a patient. FP1, FP2, (I) and (II) can be used as  
CC vaccines for inhibiting the development of cancer such as breast, ovarian  
CC or prostate cancer in a patient or a patient considered at risk for such  
CC a disease may be treated prophylactically. The fusion protein is useful  
CC for removing tumor cells from a biological sample, and for stimulating  
CC and/or expanding T cells specific for a lipophilin complex. The fusion  
CC proteins are useful as markers for progression of cancer. (II) or (III)  
CC are also useful for monitoring the progression of cancer in a patient.  
CC The present sequence represents a human lipophilin B protein sequence,  
CC which is given in an example from the present invention  
XX  
SQ Sequence 90 AA;

Query Match 100.0%; Score 450; DB 6; Length 90;  
Best Local Similarity 100.0%; Pred. No. 7e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAVKL 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAVKL 60  
QY 61 GVKRCTDQMSLQKRSLLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLLIAEVLVKILKCSV 90

RESULT 18  
ABG73168  
ID ABG73168 standard; protein; 90 AA.  
XX  
AC ABG73168;  
XX  
DT 08-APR-2003 (first entry)  
XX  
DE Human endometrial specific steroid-binding factor II (hESF II) protein.  
XX  
DE Human; endometrial specific steroid-binding factor; hESF;  
KW Clara cell 10 kDa; CC10; secretory protein; asthma;  
KW prostatic steroid-binding protein; hormone; lung; uterus; gene therapy.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= Signal\_peptide  
FT Protein 22..90  
FT /label= Mature\_hESF\_II  
XX  
PN US2002151012-A1.  
XX  
PD 17-OCT-2002.  
XX  
PF 06-NOV-2001; 2001US-00985911.  
XX  
PP 21-MAR-1996; 96US-0014724P.  
PR 21-MAR-1997; 97US-00821451.  
PR

PR 08-MAR-1999; 99US-00263810.  
PR 30-MAY-2000; 2000US-00583169.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ni J, Yu G, Gentz R;  
XX  
DR WPI; 2003-182506/18.  
DR N-PSDB; ABX14916.  
XX  
PT New human endometrial specific steroid-binding factor (hESF) proteins and  
PT genes, useful for treating or diagnosing a disease or susceptibility to a  
PT disease, particularly asthma.  
XX  
PS Claim 15; Fig 2; 37pp; English.  
XX  
CC The invention discloses isolated polypeptides, which comprise human  
CC endometrial specific steroid-binding factors I, II and III (hESF I, II  
CC and III), and the nucleic acids encoding them. The hESF polypeptide has  
CC homologs to mammalian Clara cell 10 kDa (CC10) secretory protein and  
CC rat prostatic steroid-binding protein which are factors which modulate or  
CC mediate the action of hormones involved in the regulation of functions of  
CC the lung and uterus. The nucleic acids and polypeptides can be used to  
CC identify compounds that bind to and inhibit activation, raise antibodies  
CC or develop antagonists against the isolated hESF polypeptide. The  
CC polypeptides or polynucleotides are useful for treating a patient having  
CC a need of hESF I, hESF II, hESF III or for treating a patient having a  
CC need to inhibit hESF. The polypeptide is administered by providing to the  
CC patient the DNA encoding the hESF polypeptide in vivo (gene therapy). In  
CC particular, the disease is asthma. The hESF polypeptides or  
CC polynucleotides are also useful for diagnosing a disease or a  
CC susceptibility to the disease. The sequence presented is the hESF II  
CC protein  
XX  
SQ Sequence 90 AA;

Query Match 100.0%; Score 450; DB 6; Length 90;  
Best Local Similarity 100.0%; Pred. No. 7e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAVKL 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAVKL 60  
QY 61 GVKRCTDQMSLQKRSLLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLLIAEVLVKILKCSV 90

RESULT 19  
ABR47518  
ID ABR47518 standard; protein; 90 AA.  
XX  
AC ABR47518;  
XX  
DT 12-JUN-2003 (first entry)  
XX  
DE Breast cancer associated protein sequence SEQ ID NO:272.  
XX  
KW Human; breast cancer; cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO2003004989-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 21-JUN-2002; 2002WO-US019669.  
XX  
PP 21-JUN-2001; 2001US-0299887P.  
PR 27-JUN-2001; 2001US-0301572P.  
PR 18-JUL-2001; 2001US-0306501P.  
PR 25-SEP-2001; 2001US-0325002P.  
PR



```
PR 05-MAR-2002; 2002US-0362585P.
PR 14-MAY-2002; 2002US-0380391P.
XX
XX (MILL-) MILLENIUM PHARM INC.
XX
XX Lillie J, Gannavarapu M, Glatt K, Hoerh S, Kamatkar S,
PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
PI East RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
XX
XX WPI; 2003-210381/20.
DR N-PSDB; ACC50214.
DR
XX Breast cancer diagnosis or treatment by comparing the level of expression
PT of a marker in a patient sample with that in the control non-breast
PT cancer sample.
XX
XX Claim 1; SEQ ID NO 272; 128pp; English.
XX
XX The present invention describes a method for assessing whether a patient
CC is afflicted with breast cancer. The method comprises comparing the level
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences from
CC the present invention have cytostatic activities and can be used in gene
CC therapy. The method is useful for diagnosing and treating breast cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 90 AA;
SQ
Query Match 100.0%; Score 450; DB 6; Length 90;
Best Local Similarity 100.0%; Pred. No. 7e-50;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFIISPLFKLSLAKFDAPPEAVAL 60
DB 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFIISPLFKLSLAKFDAPPEAVAL 60
QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
DB 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
RESULT 20
ADA43319
ID ADA43319 standard; protein; 90 AA.
XX
XX ADA43319;
XX
XX 20-NOV-2003 (first entry)
XX Human BUI01 polypeptide.
XX
XX Human; breast cancer; mammaglobin; BUI01; BS106.
XX
XX Homo sapiens.
XX
XX US2003044859-A1.
XX
XX 06-MAR-2003.
XX
XX 11-OCT-2001; 2001US-00975502.
XX
XX 19-AUG-1996; 96US-00697105.
XX 19-AUG-1996; 96US-00697106.
XX 15-AUG-1997; 97US-00912149.
XX 15-AUG-1997; 97US-00912149.
XX 31-OCT-1997; 97US-00962094.
XX 18-DEC-1998; 98US-00215818.
PR 05-MAR-2002; 99US-00467602.
PR 29-FEB-2000; 2000US-00516444.
XX
XX (HENS/) HENSLEE J G.
XX (FRIE/) FRIEDMAN P N.
XX
XX Henslee JG, Friedman PN;
XX
XX WPI; 2003-615776/58.
XX
XX Assay for detecting breast cancer, comprises detecting mammaglobin, BUI02
PT and/or BS106 breast cancer markers.
XX
XX Disclosure; Fig 16; 59pp; English.
XX
XX The invention relates to an assay which employs the use of two or more
CC breast cancer markers, selected from mammaglobin, BUI01 and BS106, used
CC for the detection of breast cancer in a patient. The method is used to
CC detect breast cancer by contacting a test sample from the patient with at
CC least two of the markers or at least two antibodies specific for BS106,
CC mammaglobin, BUI01 or a multimeric antigen, and correlating the presence
CC of the polypeptides with breast cancer. This sequence represents the
CC BUI01 polypeptide of the invention.
XX
XX Sequence 90 AA;
SQ
Query Match 100.0%; Score 450; DB 6; Length 90;
Best Local Similarity 100.0%; Pred. No. 7e-50;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFIISPLFKLSLAKFDAPPEAVAL 60
DB 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDFFFIISPLFKLSLAKFDAPPEAVAL 60
QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
DB 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
RESULT 21
AA024005
ID AA024005 standard; protein; 90 AA.
XX
XX AA024005;
XX
XX 06-NOV-2003 (first entry)
XX
XX Human BUI01 protein.
XX
XX Human; multimeric polypeptide antigen; MPA; BUI01; Mammaglobin;
XX uteroglobin; cytostatic; immune response inducer; cancer; fibroadenoma;
XX atypical hyperplasia; cystic breast disease.
XX
XX Homo sapiens.
XX
XX US6552164-B1.
XX
XX 22-APR-2003.
XX
XX 20-DEC-1999; 99US-00467602.
XX
XX 19-AUG-1996; 96US-00679105.
XX 19-AUG-1996; 96US-00697106.
XX 15-AUG-1997; 97US-00912149.
XX 15-AUG-1997; 97US-00912149.
XX 18-DEC-1998; 98US-00215818.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Colpitts TL, Russell JC;
XX
XX WPI; 2003-584354/55.
XX
XX N-PSDB; AAL55893.
XX
```



Db 1 MKLSVCLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAVAKL 60  
Qy 61 GVKRCTDQMSLOKRSLLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLOKRSLLIAEVLVKILKCSV 90

RESULT 24  
ADR46900  
ID ADR46900 standard; protein; 90 AA.  
XX  
AC ADR46900;  
XX  
XX 21-OCT-2004 (first entry)  
XX  
DE Breast cancer marker gene BU101 EST contig open reading frame 1.  
XX  
XX Human; Breast cancer; BU101; SNP; single nucleotide polymorphism.  
XX  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 53  
FT /note= "May be Leu as the result of a single nucleotide  
FT polymorphism"  
XX  
XX US6770435-B1.  
XX  
XX 03-AUG-2004.  
XX  
XX 01-NOV-1999; 99US-00431384.  
XX  
XX 19-AUG-1996; 96US-00697105.  
XX  
XX 15-AUG-1997; 97US-00912276.  
XX  
XX 19-JAN-1999; 99US-00233693.  
XX  
XX (ABBO ) ABBOTT LAB.  
XX  
XX Billigal PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;  
XX Granados EN, Hodges SC, Klass MR, Kratochvil JD, Russell JC;  
XX Scheffel CP, Stroupe SD;  
XX  
XX WPI; 2004-660350/64.  
XX  
XX N-PSDB; ADR46886.  
XX  
XX Detecting presence of BU101 (derived from breast tissue) nucleotide  
XX sequence in breast tissue sample, by contacting sample with nucleic acid  
XX probe having sequence complementary to BU101, and detecting presence of  
XX hybrids.  
XX  
XX Example 1; SEQ ID NO 20; 56pp; English.  
XX  
XX The invention relates to detecting the presence of a nucleotide sequence,  
XX appearing as ADR46886 (an assembled cDNA contig for cancer marker gene  
XX BU101), in breast tissue sample, involving contacting a test sample with  
XX a nucleic acid probe having a sequence complementary to ADR46886, for  
XX hybridisation to occur between the probe and ADR46886, and detecting  
XX presence of hybrids, where presence of hybrids indicates presence of  
XX nucleotide sequence ADR46886. The method is useful for detecting the  
XX presence or absence of the nucleotide sequences comprising ADR46886, in a  
XX test breast tissue sample and for detecting the diseases associated with  
XX BU101 or diseases of the breast, such as breast cancer. The present  
XX sequence is the deduced BU101 protein.  
XX  
XX Sequence 90 AA;  
XX  
XX Query Match 100.0%; Score 450; DB 8; Length 90;  
XX Best Local Similarity 100.0%; Pred. No. 7e-50;  
XX Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX

Qy 1 MKLSVCLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAVAKL 60  
Db 1 MKLSVCLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAVAKL 60

Qy 61 GVKRCTDQMSLOKRSLLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLOKRSLLIAEVLVKILKCSV 90

RESULT 25  
ABB11907  
ID ABB11907 standard; peptide; 117 AA.  
XX  
XX ABB11907;  
XX  
XX 11-JAN-2002 (first entry)  
XX  
DE Human breast tumour-associated protein homologue, SEQ ID NO:2277.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
XX haematopoiesis regulation; tissue growth; immunomodulator; activin;  
XX inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
XX chronic inflammatory condition; proliferative retinopathy;  
XX atherosclerosis; coronary heart disease; arterial ischaemia;  
XX bone disorder; osteoporosis; vascular growth disorder;  
XX tissue regeneration; wound healing; infection; immune disorder;  
XX cell culture; drug screening; gene therapy; antiinflammatory;  
XX antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
XX cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
XX antifungal; vulnery; antiulcer.  
XX  
XX Homo sapiens.  
XX  
XX WO200157188-A2.  
XX  
XX 03-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US003800.  
XX  
XX 03-FEB-2000; 2000US-00496914.  
XX  
XX 27-APR-2000; 2000US-00560875.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-457740/49.  
XX  
XX N-PSDB; ABA09151.  
XX  
XX Human proteins and DNA encoding sequences useful for preventing, treating  
XX or ameliorating a medical condition in a mammalian subject e.g. arthritis  
XX and cancer.  
XX  
XX Claim 20; Page 275; 1963pp; English.  
XX  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
XX invention also relates to vectors and recombinant host cells comprising a  
XX nucleotide of the invention, methods of producing the novel polypeptides,  
XX antibodies against the polypeptides, methods of detecting the nucleotides  
XX or polypeptides in a sample, and methods of identifying compounds which  
XX bind to polypeptides of the invention. Although novel, many of the  
XX polypeptides of the invention have homology to known proteins, thereby  
XX giving an insight into their probable biological activities, and hence  
XX potential therapeutic applications. The polypeptides of the invention may  
XX have various activities, including cytokine, cell proliferation or cell  
XX differentiation activities; stem cell growth factor activity;  
XX haematopoiesis regulatory activity; tissue growth activity;  
XX immunomodulatory activity; activin- or inhibin-related activities;  
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or  
XX thrombolytic activities; receptor or ligand activities; or may be  
XX involved in oncogenesis, cancer cell proliferation or metastasis.  
XX Depending on their biological activities, polypeptides and nucleotides of  
XX the invention are useful for preventing, treating or ameliorating medical

CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention  
XX  
SQ Sequence 117 AA;

Query Match 100.0%; Score 450; DB 4; Length 117;  
Best Local Similarity 100.0%; Pred. No. 9.5e-50;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKLSVCLLVLTALCCYQANAFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAVAAKL 60  
Db 28 MKLSVCLLVLTALCCYQANAFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAVAAKL 87  
QY 61 GVKRCTDQMSLQKRSLIAEVLVKLKCSV 90  
Db 88 GVKRCTDQMSLQKRSLIAEVLVKLKCSV 117

Search completed: September 26, 2005, 08:25:18  
Job time : 106.857 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:18:37 ; Search time 106.286 Seconds  
(without alignments)  
338.415 Million cell updates/sec

Title: US-09-975-502a-5

Perfect score: 475

Sequence: 1 MKLLMVLMLAALSLQHCYAGS.....LSNVEVFMQLIYDSSLCDLP 93

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: Geneseqp16Dec04:\*  
2: Geneseqp1980s:\*  
3: Geneseqp1990s:\*  
4: Geneseqp2000s:\*  
5: Geneseqp2001s:\*  
6: Geneseqp2002s:\*  
7: Geneseqp2003as:\*  
8: Geneseqp2003bs:\*  
9: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	100.0	93	2	AAW10179 Mammary-s
2	475	100.0	93	2	AAW59777 Mammary-s
3	475	100.0	93	2	AAW48432 Mamaglob
4	475	100.0	93	2	AAW01718 Mamaglob
5	475	100.0	93	3	AAW84622 Mamaglob
6	475	100.0	93	3	AAW13786 Human mam
7	475	100.0	93	4	AAW51127 Human mam
8	475	100.0	93	4	AAW07517 Human mam
9	475	100.0	93	5	ABG94665 Human mam
10	475	100.0	93	5	ABG78934 Human mam
11	475	100.0	93	6	ABJ37757 Human mam
12	475	100.0	93	6	AAO19897 Human mam
13	475	100.0	93	6	ABP96088 Human mam
14	475	100.0	93	6	ABP96096 Human mam
15	475	100.0	93	6	ADA83774 Human mam
16	475	100.0	93	6	ABR47521 Breast ca
17	475	100.0	93	7	AAO24004 Human mam
18	475	100.0	93	7	ABW00048 Human mam
19	475	100.0	93	7	ADL93165 Human bre
20	475	100.0	93	8	ADN04238 Antipocri
21	475	100.0	93	8	ADQ17425 Human sof
22	475	100.0	93	8	ABM81260 Tumour-as
23	475	100.0	101	5	ABG94688 Human mam
24	475	100.0	102	5	ABG94689 Human mam
25	475	100.0	132	5	ABG94684 Human mam

26	475	100.0	132	7	ABW00067	Abw00067 Ral2 (s) -
27	475	100.0	182	6	ABP96109	Abp96109 Human mam
28	475	100.0	410	4	AAU33359	Aau33359 Human bre
29	475	100.0	410	5	ABG78926	Abg78926 Human bre
30	475	100.0	410	6	ABJ37749	Abj37749 Human tum
31	475	100.0	410	7	ADL93157	Adl93157 Human bre
32	475	100.0	477	6	ABP96114	Abp96114 Mamaglob
33	475	100.0	743	4	AAU33358	Aau33358 Human bre
34	475	100.0	743	5	ABG78925	Abg78925 Human bre
35	475	100.0	743	6	ABJ37748	Abj37748 Human tum
36	475	100.0	743	7	ADL93156	Adl93156 Human bre
37	475	100.0	1095	4	AAU33357	Aau33357 Human bre
38	475	100.0	1095	5	ABG78924	Abg78924 Human bre
39	475	100.0	1095	6	ABJ37747	Abj37747 Human tum
40	475	100.0	1095	7	ADL93155	Adl93155 Human bre
41	474	99.8	93	4	AAE07531	Aae07531 Human mam
42	474	99.8	93	6	ABP96092	Abp96092 Human mam
43	471	99.2	93	4	AAE07534	Aae07534 Human mam
44	471	99.2	93	4	AAE07536	Aae07536 Human mam
45	471	99.2	93	4	AAE07535	Aae07535 Human mam
46	471	99.2	93	6	ABP96107	Abp96107 Human mam
47	471	99.2	93	6	ABP96106	Abp96106 Human mam
48	471	99.2	93	6	ABP96105	Abp96105 Human mam
49	471	99.2	182	6	ABP96110	Abp96110 Human mam
50	468	98.5	93	4	AAE07529	Aae07529 Human mam
51	468	98.5	93	6	ABP96091	Abp96091 Human mam
52	466	98.1	93	4	AAE07532	Aae07532 Human mam
53	466	98.1	93	6	ABP96093	Abp96093 Human mam
54	463	97.5	93	4	AAE07533	Aae07533 Human mam
55	463	97.5	93	4	AAE07530	Aae07530 Human mam
56	463	97.5	93	6	ABP96095	Abp96095 Human mam
57	463	97.5	93	6	ABP96094	Abp96094 Human mam
58	448.5	94.4	90	4	AAE07528	Aae07528 Human mam
59	448.5	94.4	90	4	AAE07537	Aae07537 Human mam
60	448.5	94.4	90	6	ABP96090	Abp96090 Human mam
61	448.5	94.4	90	6	ABP96108	Abp96108 Human mam
62	414	87.2	80	5	ABG94687	Abg94687 Human mam
63	414	87.2	95	5	ABG94686	Abg94686 Human mam
64	399	84.0	145	6	ABP96112	Abp96112 Human mam
65	386	81.3	76	5	ABG94690	Abg94690 Human mam
66	386	81.3	145	6	ABP96111	Abp96111 Human mam
67	386	81.3	460	6	ABP96113	Abp96113 Mamaglob
68	380	80.0	74	3	AAW84624	Aaw84624 Amino aci
69	380	80.0	220	4	AAO22141	Aao22141 Ral2-mamm
70	281	59.2	95	2	AAW35804	Aaw35804 Human end
71	281	59.2	95	2	AAW02590	Aaw02590 A human m
72	281	59.2	95	3	AAW03769	Aaw03769 Human end
73	281	59.2	95	3	AAW65394	Aaw65394 Human 5'
74	281	59.2	95	3	AAW92226	Aaw92226 Human end
75	281	59.2	95	3	AAW92237	Aaw92237 Mamaglob
76	281	59.2	95	4	AAW31682	Aaw31682 An endome
77	281	59.2	95	5	AAO20555	Aao20555 Protein o
78	281	59.2	95	5	ABG96366	Abg96366 Human ova
79	281	59.2	95	5	ABW09635	Abw09635 Human end
80	281	59.2	95	6	ABG73169	Abg73169 Human end
81	281	59.2	95	6	ABR47522	AbR47522 Breast ca
82	281	59.2	95	7	ADK41999	Adk41999 Human PRO
83	281	59.2	95	7	ADK41999	Adk41999 Human MAM
84	281	59.2	95	7	ADN39186	Adn39186 Cancer/an
85	281	59.2	95	8	ADQ18632	Adq18632 Human sof
86	281	59.2	108	2	AAW60038	Aaw60038 Human end
87	272	57.3	93	5	AAU83620	Aau83620 Human PRO
88	272	57.3	93	6	ABO33733	AbO33733 Novel hum
89	272	57.3	93	6	ABU82076	Abu82076 Novel hum
90	272	57.3	93	6	ABJ72256	Abj72256 Human PRO
91	272	57.3	93	6	ABJ72384	Abj72384 Human PRO
92	272	57.3	93	6	ABO34279	AbO34279 Human sec
93	272	57.3	93	7	ABJ72086	Abj72086 Human mem
94	272	57.3	93	7	ADB83548	AdB83548 Novel hum
95	272	57.3	93	7	ADB80654	AdB80654 Novel hum
96	272	57.3	93	7	ADB73195	AdB73195 Novel hum
97	272	57.3	93	7	ADB78277	AdB78277 Novel hum
98	272	57.3	93	7	ADB78277	AdB78277 Novel hum



```

XX 13-JUL-1998 (first entry)
DT Mammaglobin protein.
DE Mammaglobin; detection; diagnosis; breast cancer; tumour; antibody;
KW gene therapy; human.
XX Homo sapiens.
XX WO9807753-A1.
XX 26-FEB-1998.
XX 19-AUG-1997; 97WO-US014666.
XX 19-AUG-1996; 96US-00697106.
XX 15-AUG-1997; 97US-00912149.
XX (ABBO ) ABBOTT LAB.
XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;
PI Granados EN, Hodges SC, Klaas MR, Kratochvil JD, Roberts-Rapp L;
PI Russell JC, Stroupe SD;
XX WPI; 1998-169096/15.
DR N-PSDB; AAV17905; AAV17906.
XX Antibodies to mammaglobin polypeptide(s) - used for detecting,
PT diagnosing, preventing or treating diseases or conditions of breast such
PT as breast cancer.
XX Claim 8; Page 92; 105pp; English.
XX The present sequence represents mammaglobin which is used in an example
PS of the present invention. The present invention describes an antibody (A)
CC which specifically binds to at least 1 mammaglobin epitope (ME) which is
CC derived from an amino acid sequence having at least 50% identity to an
CC amino acid sequence (see AAW48432) and fragments. Also described are: (1)
CC an assay kit for determining the presence of mammaglobin antigen (MA) in
CC a test sample, comprising a container containing an antibody as in (A);
CC (2) a method for producing antibodies which specifically bind to a MA,
CC comprising administering an isolated immunogenic polypeptide or fragment
CC to elicit an immune response, where the immunogenic polypeptide comprises
CC at least 1 ME and has at least 50% identity to a sequence (see AAW48432)
CC and fragments, and (3) a method for producing antibodies which
CC specifically bind to a MA comprising administering to a mammal a plasmid
CC comprising a sequence which encodes at least 1 ME derived from a
CC polypeptide having an amino acid sequence (see AAW48432) and fragments.
CC The products and methods can be used for detecting, diagnosing, staging,
CC monitoring, prognosticating, preventing or treating, or determining
CC predisposition to diseases or conditions of the breast such as breast
CC cancer.
XX Query Match 100.0%; Score 475; DB 2; Length 93;
XX Best Local Similarity 100.0%; Pred. No. 6.1e-47;
XX Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKLLMVLMLAALSQHCYAGSGCPLENNVISKTNIPQVSKTEYKELQEFIDNATTNAID 60
Db 1 MKLLMVLMLAALSQHCYAGSGCPLENNVISKTNIPQVSKTEYKELQEFIDNATTNAID 60
Qy 61 ELKCEFLNQDTETLSNVEVFQMQLIYDSSLCDLF 93
Db 61 ELKCEFLNQDTETLSNVEVFQMQLIYDSSLCDLF 93
RESULT 5
AAV84622
ID AAY84622 standard; protein; 93 AA.
XX AC AAY84622;
XX 25-JUL-2000 (first entry)
DT Amino acid sequence of the mammary-specific protein mammaglobin.
DE Human; mammaglobin; mammary gland; breast cancer; endometrial cancer.
XX Homo sapiens.
XX WO200018783-A1.

```

---

```

AC AAY01718;
XX 25-JUN-1999 (first entry)
XX Mammaglobin, a mammary specific protein.
XX Human; mammary-specific protein; mammaglobin; antigen; vaccine;
KW mammaglobin-expressing cancer; breast cancer;
XX autologous tumor lymphocyte; diagnosis; marker.
XX OS Homo sapiens.
XX WO9914230-A1.
XX 25-MAR-1999.
XX 18-SEP-1998; 98WO-US017991.
XX 18-SEP-1997; 97US-00933149.
XX (UNIW ) UNIV WASHINGTON.
XX Watson MA, Fleming TP;
XX WPI; 1999-244021/20.
XX N-PSDB; AAX26966.
XX Mammaglobin, secreted protein overexpressed in breast cancer.
XX Claim 15; Fig 2; 60pp; English.
XX The present sequence represents a human mammary-specific protein,
CC designated mammaglobin. The specification describes a protein comprising
CC a mammaglobin antigen that is recognized by B and/or Tc cells specific
CC for the natural, secreted and glycosylated form of mammaglobin
CC polypeptide. This protein, or recombinant vectors that express it, are
CC used in vaccines for treating mammaglobin-expressing cancers.
CC specifically of the breast. Such cancers can also be treated using
CC autologous tumor lymphocytes activated ex vivo with an mammaglobin
CC antigen, then returned to the patient. Expression of mammaglobin is
CC elevated in 27% of stage I primary breast cancers, so it represents a
CC marker useful for diagnosis of this disease
XX Sequence 93 AA;
XX Query Match 100.0%; Score 475; DB 2; Length 93;
XX Best Local Similarity 100.0%; Pred. No. 6.1e-47;
XX Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKLLMVLMLAALSQHCYAGSGCPLENNVISKTNIPQVSKTEYKELQEFIDNATTNAID 60
Db 1 MKLLMVLMLAALSQHCYAGSGCPLENNVISKTNIPQVSKTEYKELQEFIDNATTNAID 60
Qy 61 ELKCEFLNQDTETLSNVEVFQMQLIYDSSLCDLF 93
Db 61 ELKCEFLNQDTETLSNVEVFQMQLIYDSSLCDLF 93
RESULT 5
AAV84622
ID AAY84622 standard; protein; 93 AA.
XX AC AAY84622;
XX 25-JUL-2000 (first entry)
DT Amino acid sequence of the mammary-specific protein mammaglobin.
DE Human; mammaglobin; mammary gland; breast cancer; endometrial cancer.
XX Homo sapiens.
XX WO200018783-A1.

```

```

XX PD 06-APR-2000.
XX PF 29-SEP-1999; 99WO-US022616.
XX PR 29-SEP-1998; 98US-00162622.
XX PA (UNIW ) UNIV WASHINGTON.
XX PI Watson MA, Fleming TP;
XX PI WPI; 2000-293105/25.
XX DR N-PSDB; AAA12632.
XX DR
XX PT Methods for detecting breast cancer, comprising detecting elevated
XX PT concentrations of a mammaglobin polypeptide, using an antibody, or
XX PT detecting elevated concentrations of the mRNA encoding the polypeptide,
XX PT using oligonucleotides.
XX PS Example 1; Fig 2; 71pp; English.
XX CC The present sequence represents the human mammary-specific secreted
XX CC protein mammaglobin. Mammaglobin expression is restricted to the mammary
XX CC gland. Dysregulation of the mammaglobin gene occurs early and frequently
XX CC in breast cancer. The specification describes a method for detecting the
XX CC presence of breast cancer in a patient, comprising detecting an elevated
XX CC concentration of mRNA encoding a mammaglobin polypeptide. The methods are
XX CC useful for detecting the presence of breast and endometrial cancer
XX CC
XX CC Sequence 93 AA;
XX CC
XX CC Query Match 100.0%; Score 475; DB 3; Length 93;
XX CC Best Local Similarity 100.0%; Pred. No. 6.1e-47;
XX CC Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX CC
QY 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDNATTNAID 60
DB 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDNATTNAID 60
QY 61 ELKECFNLQTDLTLSNVEVFQMQLIYDSSLCDLF 93
DB 61 ELKECFNLQTDLTLSNVEVFQMQLIYDSSLCDLF 93

RESULT 6
AAB13786
ID AAB13786 standard; protein; 93 AA.
XX AC AAB13786;
XX DT 20-JUN-2001 (first entry)
XX DE Human mammaglobin.
XX KW Human; breast cancer; breast disease detection; mammaglobin; uteroglobin;
XX KW chromosome 11q13; BU101; endometrial; cytostatic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Modified-site 53
XX FT Modified-site 68 /note= "Optionally N-glycosylated"
XX FT Modified-site 68 /note= "Optionally N-glycosylated"
XX PN WC200035950-A2.
XX PD 22-JUN-2000.
XX PF 20-DEC-1999; 99WO-US030489.
XX PR 18-DEC-1998; 98US-00215818.
XX

```

```

PA (ABBO ) ABBOTT LAB.
XX Colpitts TL, Russell JE;
XX WPI; 2000-442366/38.
XX DR N-PSDB; AAA64845.
XX PT Multimeric polypeptide antigen and antibody specific to the antigen are
XX PT useful for diagnosing, detecting and treating breast cancer.
XX PS Claim 1; Page 123-124; 124pp; English.
XX CC Mammaglobin is a member of the uteroglobin protein family. The
XX CC mammaglobin gene has been localised to chromosome 11q13. The present
XX CC sequence is the protein sequence for human mammaglobin. The present
XX CC invention relates to a multimeric polypeptide antigen, which comprises of
XX CC the present sequence and BU101 polypeptide (AAB13787). BU101 is another
XX CC uteroglobin protein. The presence of multimeric polypeptide antigen in a
XX CC test sample can be used as the basis for a test to diagnose breast
XX CC disease e.g. breast cancer, in a patient. The detection can be carried
XX CC out using antibodies specific for the multimeric polypeptide antigen
XX CC
XX CC Sequence 93 AA;
XX CC
XX CC Query Match 100.0%; Score 475; DB 3; Length 93;
XX CC Best Local Similarity 100.0%; Pred. No. 6.1e-47;
XX CC Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX CC
QY 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDNATTNAID 60
DB 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDNATTNAID 60
QY 61 ELKECFNLQTDLTLSNVEVFQMQLIYDSSLCDLF 93
DB 61 ELKECFNLQTDLTLSNVEVFQMQLIYDSSLCDLF 93

RESULT 7
AAB51127
ID AAB51127 standard; protein; 93 AA.
XX AC AAB51127;
XX DT 20-MAR-2001 (first entry)
XX DE Human mammaglobin amino acid sequence SEQ ID NO:27.
XX KW Human; mammaglobin; breast cancer; detection; diagnosis; antibody;
XX KW vaccine; cytostatic; antimammaglobin.
XX OS Homo sapiens.
XX PN WO2000073338-A1.
XX PD 07-DEC-2000.
XX PF 26-MAY-2000; 2000WO-US014845.
XX PR 28-MAY-1999; 99US-0136528P.
XX PR 01-JUN-1999; 99US-0137048P.
XX PA (CORI-) CORIXA CORP.
XX PA (HEND/) HENDRICKSON R C.
XX PA (HOUG/) HOUGHTON R L.
XX PA (REED/) REED S G.
XX PI Fanger GR;
XX WPI; 2001-049928/06.
XX PT Polypeptide comprising at least seven consecutive amino acid residues of
XX PT human mammaglobin, useful in the treatment and detection of breast
XX PT cancer.

```



XX Example 1; Fig 2; 109pp; English.

XX The present invention describes human mammaglobin peptides (I) comprising

CC at least 7 consecutive residues. Also described are: (1) a vaccine

CC comprising (I) with an immunostimulant which is an adjuvant; (2) an

CC isolated antibody (Ab1) or its antigen-binding fragment, which

CC specifically binds to a mammaglobin epitope having the sequence of Pro2-3

CC; (3) an isolated antibody (Ab2) or its antigen-binding fragment that

CC specifically binds to glycosylated mammaglobin; (4) a method for

CC inhibiting the development of breast cancer in a patient, comprising

CC administering (I) or Ab1 or Ab2; and (5) a method (M1) for determining

CC the presence or absence of breast cancer in a patient. (I) has cytostatic

CC activity. The polypeptides and antibodies are used in vaccines for the

CC prevention and treatment of breast cancer. They are also used for

CC diagnosis and monitoring of breast cancer. The present sequence

CC represents the human mammaglobin amino acid sequence, which is used in an

XX example from the present invention

SQ Sequence 93 AA;

Query Match 100.0%; Score 475; DB 4; Length 93;

Best Local Similarity 100.0%; Pred. No. 6.1e-47;

Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSOHCYAGSGCPLENNVISKTNPOVSKTEYKELQEFIDDDNATTNAID 60

DB 1 MKLLMVLMLAALSOHCYAGSGCPLENNVISKTNPOVSKTEYKELQEFIDDDNATTNAID 60

QY 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93

DB 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93

RESULT 8

AAE07517

ID AAE07517 standard; protein; 93 AA.

XX AAE07517;

DT 06-NOV-2001 (first entry)

XX Human mammaglobin protein.

XX Human; mammaglobin; lipophilin; cytostatic; vaccine; gene therapy;

KW uteroglobin; cancer; breast; ovary; prostate.

XX Homo sapiens.

XX WO200158947-A1.

PN 16-AUG-2001.

XX 08-FEB-2001; 2001WO-US0004439.

XX 11-FEB-2000; 2000US-0183495P.

PR 28-JUN-2000; 2000US-0215735P.

XX (CORI-) CORIXA CORP.

XX Carter D, Vedvick TS, Vallieve-Douglass J, Houghton RL, Dillon DC;

PI WPI; 2001-497069/54.

XX N-PSDB; AAD13755.

XX Novel isolated complex two lipophilin-like polypeptides linked by at

PT least one disulfide bond, used to treat or prevent breast, ovarian or

PT prostate cancer.

XX Example 5; Page 72; 91pp; English.

PS The invention relates to a complex comprising a lipophilin-like

XX polypeptide linked by at least one disulfide bond to a second lipophilin

CC -like polypeptide. Lipophilin-like protein are members of uteroglobin

CC superfamily. Lipophilin-like proteins are useful in the preparation of

CC vaccines. The complex containing lipophilin-like proteins are useful for

CC treating or preventing breast, ovarian or prostate cancer. The complex is

CC also used for determining the presence or absence of cancer in a patient,

CC or monitor the progress of cancer in a patient. Lipophilin DNA is also

CC useful in gene therapy. The present sequence is human mammaglobin which

CC is a lipophilin-like protein

XX Sequence 93 AA;

Query Match 100.0%; Score 475; DB 4; Length 93;

Best Local Similarity 100.0%; Pred. No. 6.1e-47;

Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSOHCYAGSGCPLENNVISKTNPOVSKTEYKELQEFIDDDNATTNAID 60

DB 1 MKLLMVLMLAALSOHCYAGSGCPLENNVISKTNPOVSKTEYKELQEFIDDDNATTNAID 60

QY 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93

DB 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93

RESULT 9

ABG94665

ID ABG94665 standard; protein; 93 AA.

XX ABG94665;

DT 02-DEC-2002 (first entry)

XX Human mammaglobin protein sequence.

XX Human; epitope; mammaglobin; breast cancer; cytostatic; T cell; CD4+;

KW CD8+; antigen.

XX Homo sapiens.

XX WO200253017-A2.

PN 11-JUL-2002.

XX 08-JAN-2002; 2002WO-US003057.

XX 08-JAN-2001; 2001US-00757417.

PR 08-NOV-2001; 2001US-00008045.

XX (CORI-) CORIXA CORP.

XX Fling SP, Foy TM, Clapper JD, Wang A, Johnson JC, Mcneill PD;

PI Sutherland RA;

XX WPI; 2002-706844/76.

DR N-PSDB; ABS71817.

XX Novel polypeptides comprise one or more human mammaglobin epitopes and

PT polynucleotides encoding the polypeptides, useful for preventing and

PT treating breast cancers.

XX Claim 1; Fig 2; 121pp; English.

PS The invention relates to an isolated polypeptide (I) comprising 7-30

XX consecutive amino acid residues of human mammaglobin, where one or more

CC mammaglobin-specific T cells specifically reacts with (I). Also included

CC are (1) a composition comprising (I), in combination with a

CC physiologically acceptable carrier or immunostimulant; (2) a diagnostic

CC kit, comprising (I) and a detection reagent comprising a reporter group;

CC (3) removing tumour cells from a biological sample, by contacting a

CC biological sample with T cells that specifically react with (I), under

CC conditions and for a time sufficient to permit the removal of cells

CC expressing mammaglobin or a peptide epitope from the sample; (4) an

CC isolated T cell population (II), comprising T cells prepared using (I);

CC and (5) a composition (C2) comprising a polynucleotide comprising a  
 CC sequence encoding human mammaglobin (or variant or tagged with an  
 CC affinity tag), or a polypeptide comprising a human mammaglobin (or  
 CC variant or tagged with an affinity tag), in combination with an  
 CC immunostimulant. The peptide is useful for inhibiting the development of  
 CC breast cancer in a patient, by incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells  
 CC isolated from a patient with a mammaglobin epitope, such that T cells  
 CC proliferate, administering the proliferated T cells to the patient,  
 CC optionally cloning at least one proliferated T cell and administering at  
 CC least one cloned cell, and thus inhibiting the development of breast  
 CC cancer in the patient. The compositions and methods are useful for  
 CC inhibiting the development of breast cancer in a patient. The present  
 CC sequence is human mammaglobin

XX  
 SQ Sequence 93 AA;

Query Match 100.0%; Score 475; DB 5; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-47;  
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSQHCVAGSGCPLEENVISKTINPQVSKTEYKELLQEFIDNATTNAID 60  
 DB 1 MKLLMVLMLAALSQHCVAGSGCPLEENVISKTINPQVSKTEYKELLQEFIDNATTNAID 60

QY 61 ELKECFLNQDTETLSNVFVMQLIYDSSICDLF 93  
 DB 61 ELKECFLNQDTETLSNVFVMQLIYDSSICDLF 93

RESULT 10  
 ABG78934  
 ID ABG78934 standard; protein; 93 AA.  
 AC ABG78934;  
 DT 15-NOV-2002 (first entry)  
 DE Human breast tumour polypeptide #25.  
 DE Human; breast tumour protein; breast cancer; cytostatic; vaccine.  
 KW Homo sapiens.  
 OS Homo sapiens.  
 XX US2002085998-A1.  
 XX 04-JUL-2002.  
 XX 13-APR-2001; 2001US-00834759.  
 XX 28-DEC-1998; 98US-00222575.  
 XX 02-APR-1999; 99US-00285480.  
 XX 23-JUN-1999; 99US-00339338.  
 XX 02-SEP-1999; 99US-00389681.  
 XX 03-NOV-1999; 99US-00433826.  
 XX 17-APR-2000; 2000US-00551621.  
 XX 08-JUN-2000; 2000US-00590751.  
 XX 22-JUN-2000; 2000US-00604287.  
 XX 20-JUL-2000; 2000US-00620405.  
 XX (CORI-) CORIXA CORP.  
 PA Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;  
 PI Henderson RA;  
 XX WPI; 2002-635657/68.  
 XX Novel breast cancer polynucleotides and polypeptides encoded by the  
 PT polynucleotides, useful for detecting the presence of breast cancer in a  
 PT patient, and in pharmaceutical compositions, for treating breast cancer.  
 XX Claim 8; Page 230; 247pp; English.  
 XX The invention relates to an isolated breast tumour polynucleotide and the

CC polypeptide it encodes. The polynucleotide and polypeptide are useful for  
 CC detecting the presence of breast cancer in a patient, and in  
 CC pharmaceutical compositions for treating breast cancer. The sequences are  
 CC useful for stimulating an immune response in a patient and can therefore  
 CC be used in production of vaccines. The sequences are also useful for  
 CC detecting the presence of a cancer in a patient, by obtaining a  
 CC biological sample from the patient, contacting the amount of polynucleotide  
 CC a composition of the invention and detecting the amount of polynucleotide  
 CC that hybridizes to the sample. This sequence represents a human breast  
 CC tumour polypeptide of the invention

XX  
 SQ Sequence 93 AA;

Query Match 100.0%; Score 475; DB 5; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-47;  
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSQHCVAGSGCPLEENVISKTINPQVSKTEYKELLQEFIDNATTNAID 60  
 DB 1 MKLLMVLMLAALSQHCVAGSGCPLEENVISKTINPQVSKTEYKELLQEFIDNATTNAID 60

QY 61 ELKECFLNQDTETLSNVFVMQLIYDSSICDLF 93  
 DB 61 ELKECFLNQDTETLSNVFVMQLIYDSSICDLF 93

RESULT 11  
 ABJ37757  
 ID ABJ37757 standard; protein; 93 AA.  
 AC ABJ37757;  
 DT 15-MAY-2003 (first entry)  
 DE Human tumour-related protein - SEQ ID No 503.  
 KW Human; vaccine; gene therapy; T cell stimulation; T cell expansion;  
 KW tumour; breast cancer; cancer; immune response stimulation.  
 XX Homo sapiens.  
 XX WO200283956-A1.  
 XX 24-OCT-2002.  
 XX 15-APR-2002; 2002WO-US012378.  
 XX 13-APR-2001; 2001US-00834759.  
 XX 07-DEC-2001; 2001US-00007805.  
 XX 13-FEB-2002; 2002US-00076622.  
 XX (CORI-) CORIXA CORP.  
 XX Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;  
 PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;  
 PI Vedwick TS, McNeill PD, Durham M;  
 XX WPI; 2003-103376/09.  
 XX New polypeptide and polynucleotide useful for stimulating and/or  
 PT expanding T cells specific for a tumor protein and treating breast  
 PT cancer.  
 XX Disclosure; Page 316; 375pp; English.  
 XX The invention comprises a method of stimulating and/or expanding T cells  
 CC specific for a tumour protein. The invention further comprises human  
 CC nucleic acids and proteins that are associated with tumours (e.g. breast  
 CC cancer). The method and sequences of the invention are useful for  
 CC stimulating and/or expanding T cells specific for a tumour protein,  
 CC detecting the presence of cancer, stimulating an immune response in a  
 CC patient and treating breast cancer. The present amino acid sequence  
 CC represents a human tumour-related protein

```
XX SQ Sequence 93 AA;
Query Match 100.0%; Score 475; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 6.1e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLVLMALALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNNATTNAID 60
DB 1 MKLLVLMALALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNNATTNAID 60

QY 61 ELKECFNLQTDETLSNVEVFQMQLIYDSSLCDLF 93
DB 61 ELKECFNLQTDETLSNVEVFQMQLIYDSSLCDLF 93

RESULT 12
AAO19897
ID AAO19897 standard; protein; 93 AA.
AC AAO19897;
XX
XX 11-AUG-2003 (first entry)
DT
DE Human mammaglobin A.
XX
XX UGRP1; human; mouse; promoter; uteroglobin related protein 1;
XX respiratory disorder; asthma.
XX
XX Homo sapiens.
XX
XX WO2003000111-A2.
XX
XX 03-JAN-2003.
XX
XX 18-JUN-2002; 2002WO-US019456.
XX
XX 20-JUN-2001; 2001US-0299828P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kimura S, Niimi T;
XX
XX WPI; 2003-184004/18.
XX
XX New human UGRP1 nucleic acid, useful for diagnosing or predicting a
XX predispotion to develop a respiratory disorder or determining the
XX prognosis of a subject having or suspected of having a respiratory
XX disorder e.g., asthma.
XX
XX Disclosure; Page 81; 83pp; English.
XX
XX The present invention provides the human and murine uteroglobin related
XX protein 1 (UGRP1) promoters. The sequences can be used in the diagnosis
XX of and prediction of predisposition to respiratory disorders such as
XX asthma. The present sequence is a protein sequence shown in the
XX exemplification of the invention
XX
XX Sequence 93 AA;
Query Match 100.0%; Score 475; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 6.1e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLVLMALALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNNATTNAID 60
DB 1 MKLLVLMALALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNNATTNAID 60

QY 61 ELKECFNLQTDETLSNVEVFQMQLIYDSSLCDLF 93
DB 61 ELKECFNLQTDETLSNVEVFQMQLIYDSSLCDLF 93
```

```
RESULT 13
ABP96088
ID ABP96088 standard; protein; 93 AA.
XX
XX AC ABP96088;
XX
XX 08-MAY-2003 (first entry)
DT
XX Human mammaglobin amino acid sequence SEQ ID NO:1.
XX
XX Human; lipophilin; mammaglobin; breast tumour antigen; fusion protein;
XX cancer; cytostatic; vaccine; antibody therapy; tumour.
XX
XX Homo sapiens.
XX
XX WO2003005888-A2.
XX
XX 23-JAN-2003.
XX
XX 11-JUL-2002; 2002WO-US022325.
XX
XX 13-JUL-2001; 2001US-00905673.
XX
XX 12-MAR-2002; 2002US-00096319.
XX
XX (CORI-) CORIXA CORP.
XX
XX Ranger CR, Durham M, Houghton RL, Dillon DC, Carter D;
XX Persing DH;
XX
XX WPI; 2003-229428/22.
XX
XX Fusion protein useful for preventing, treating breast, ovarian and
XX prostate cancer, has lipophilin-like polypeptide linked to another
XX lipophilin-like polypeptide or a breast tumor antigen by a peptide bond.
XX
XX Example 5; Page 105; 132pp; English.
XX
XX The present invention describes a fusion protein (FP1) comprising a first
XX lipophilin-like polypeptide linked by a peptide bond to a second
XX lipophilin-like polypeptide, or a fusion protein (FP2) comprising a first
XX lipophilin-like polypeptide linked by a peptide bond to a breast tumor
XX antigen. Also described: (1) a polynucleotide (I) encoding FP1 or FP2;
XX (2) an isolated antibody (II) or its antigen-binding fragment that
XX specifically binds to FP1 or FP2, and does not detectably bind to
XX mammaglobin, lipophilin A, B or C, or B30SD; and (3) an isolated antibody
XX (III) or its antigen-binding fragment that specifically binds to a
XX lipophilin complex. FP1 and FP2 have cytostatic activities and can be
XX used in vaccines and in antibody therapy. (II) specific for FP1 or (III)
XX can be used for determining the presence or absence of breast, ovarian or
XX prostate cancer in a patient. FP1, FP2, (I) and (II) can be used as
XX vaccines for inhibiting the development of cancer such as breast, ovarian
XX or prostate cancer in a patient or a patient considered at risk for such
XX a disease may be treated prophylactically. The fusion protein is useful
XX for removing tumor cells from a biological sample, and for stimulating
XX and/or expanding T cells specific for a lipophilin complex. The fusion
XX proteins are useful as markers for progression of cancer. (II) or (III)
XX are also useful for monitoring the progression of cancer in a patient.
XX The present sequence represents a human mammaglobin protein which is used
XX in an example from the present invention
XX
XX Sequence 93 AA;
Query Match 100.0%; Score 475; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 6.1e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLVLMALALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNNATTNAID 60
DB 1 MKLLVLMALALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNNATTNAID 60

QY 61 ELKECFNLQTDETLSNVEVFQMQLIYDSSLCDLF 93
DB 61 ELKECFNLQTDETLSNVEVFQMQLIYDSSLCDLF 93
```

RESULT 14	61	ELKECFLNQTDFTLSNVEVFMQLIYDSSILCDLF	93
ABP96096	61	ELKECFLNQTDFTLSNVEVFMQLIYDSSILCDLF	93
XX			
AC	ABP96096		
XX			
DT	08-MAY-2003	(first entry)	
XX			
DE	Human mamaglobin protein sequence SEQ ID NO:34.		
XX			
KW	Human; lipophilin; mamaglobin; breast tumour antigen; fusion protein;		
KW	cancer; cytostatic; vaccine; antibody therapy; tumour.		
XX			
OS	Homo sapiens.		
OS	Synthetic.		
XX			
PN	WO2003005888-A2.		
XX			
PD	23-JAN-2003.		
XX			
PF	11-JUL-2002; 2002WO-US022325.		
XX			
PR	13-JUL-2001; 2001US-00905673.		
PR	12-MAR-2002; 2002US-00096319.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Fanger GR, Durham M, Houghton RL, Dillon DC, Carter D;		
PI	Persing DH;		
XX			
XX	WPI; 2003-229428/22.		
XX			
PT	Fusion protein useful for preventing, treating breast, ovarian and		
PT	prostate cancer, has lipophilin-like polypeptide linked to another		
PT	lipophilin-like polypeptide or a breast tumor antigen by a peptide bond.		
XX			
PS	Example 5; Page 114; 132pp; English.		
XX			
CC	The present invention describes a fusion protein (FP1) comprising a first		
CC	lipophilin-like polypeptide linked by a peptide bond to a second		
CC	lipophilin-like polypeptide, or a fusion protein (FP2) comprising a first		
CC	lipophilin-like polypeptide linked by a peptide bond to a breast tumour		
CC	antigen. Also described: (1) a polynucleotide (I) encoding FP1 or FP2;		
CC	(2) an isolated antibody (II) or its antigen-binding fragment that		
CC	specifically binds to FP1 or FP2, and does not detectably bind to		
CC	mamaglobin, lipophilin A, B or C, or B305D; and (3) an isolated antibody		
CC	(III) or its antigen-binding fragment that specifically binds to a		
CC	lipophilin complex. FP1 and FP2 have cytostatic activities and can be		
CC	used in vaccines and in antibody therapy. (II) specific for FP1 or (III)		
CC	can be used for determining the presence or absence of breast, ovarian or		
CC	prostate cancer in a patient. FP1, FP2, (I) and (II) can be used as		
CC	vaccines for inhibiting the development of cancer such as breast, ovarian		
CC	or prostate cancer in a patient or a patient considered at risk for such		
CC	a disease may be treated prophylactically. The fusion protein is useful		
CC	for removing tumour cells from a biological sample, and for stimulating		
CC	and/or expanding T cells specific for a lipophilin complex. The fusion		
CC	proteins are useful as markers for progression of cancer. (II) or (III)		
CC	are also useful for monitoring the progression of cancer in a patient.		
CC	The present sequence represents a human mamaglobin protein sequence,		
CC	which is given in an example from the present invention		
XX			
SQ	Sequence 93 AA;		
Query Match	100.0%;	Score 475;	DB 6; Length 93;
Best Local Similarity	100.0%;	Pred. No. 6.1e-47;	
Matches	93;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MKLLVMMLAALSOHCYAGSGCPLLENVSKTINPOVSKTEYKELLQEFIDNATTNAID	60
DB	1	MKLLVMMLAALSOHCYAGSGCPLLENVSKTINPOVSKTEYKELLQEFIDNATTNAID	60

QY	61	ELKECFLNQTDFTLSNVEVFMQLIYDSSILCDLF	93
DB	61	ELKECFLNQTDFTLSNVEVFMQLIYDSSILCDLF	93
XX			
AC	ADA83774		
XX			
DT	20-NOV-2003	(first entry)	
XX			
DE	Human SCGB2A2 protein.		
XX			
KW	human; marker; expressed sequence tag; EST; arabidopsis; tumour;		
KW	stresses-induced phenotype; hyperosmotic stress; colon cancer; immunogen;		
KW	vaccine.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2002103028-A2.		
XX			
PD	27-DEC-2002.		
XX			
PF	30-MAY-2002; 2002WO-IB004189.		
XX			
PR	30-MAY-2001; 2001US-0293999P.		
PR	22-OCT-2001; 2001US-0330457P.		
PR	19-FEB-2002; 2002US-0357144P.		
XX			
PA	(BIOM-) BIOMEDICAL CENT.		
XX			
PI	Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;		
XX			
DR	WPI; 2003-175241/17.		
DR	N-PSDB; ADA83773.		
XX			
PT	Determining if a nucleic acid is a marker for a phenotype/cell type of		
PT	interest, by global comparison of expressed sequence tags known to be		
PT	expressed in the phenotype/cell type with all ESTs expressed in normal		
XX	tissue.		
PS	Claim 29; Page 161; 516pp; English.		
XX			
CC	The invention relates to a novel method for determining if a nucleic acid		
CC	is a marker for a predetermined phenotype/cell type of interest from a		
CC	biological species. The method comprises performing a global comparison		
CC	of a group of expressed sequence tags (ESTs) known to be expressed in the		
CC	phenotype/cell type of interest with all ESTs expressed in normal tissue		
CC	in order to identify ESTs that are preferentially expressed in the		
CC	phenotype/cell of interest. A method of the invention is useful for		
CC	determining whether a nucleic acid is a marker for a predetermined		
CC	phenotype or cell type of interest from a biological species, preferably		
CC	Arabidopsis or human. The cell type of interest is an abnormal cell such		
CC	as a tumour cell, and the predetermined phenotype is a stress-induced		
CC	phenotype such as hyperosmotic stress or high salt conditions. A method		
CC	of the invention is also useful for determining the progression of colon		
CC	cancer in a human, for detecting a tumour cell, and for regulating or		
CC	preventing the growth of a tumour cell. An antibody of the invention is		
CC	useful for detecting the absence or presence of peptides encoded by		
CC	tumour-associated markers. A polypeptide of the invention is useful as an		
CC	immunogen for vaccinating an animal. The present sequence represents a		
CC	tumour-associated antigen of the invention.		
XX			
SQ	Sequence 93 AA;		
Query Match	100.0%;	Score 475;	DB 6; Length 93;
Best Local Similarity	100.0%;	Pred. No. 6.1e-47;	
Matches	93;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MKLLVMMLAALSOHCYAGSGCPLLENVSKTINPOVSKTEYKELLQEFIDNATTNAID	60
DB	1	MKLLVMMLAALSOHCYAGSGCPLLENVSKTINPOVSKTEYKELLQEFIDNATTNAID	60



QY 61 ELKECFLNQTDETLSNVVFQMQLIYDSSLCDLF 93  
Db 61 ELKECFLNQTDETLSNVVFQMQLIYDSSLCDLF 93

RESULT 18  
ABW00048  
ID ABW00048 standard; protein; 93 AA.  
XX  
AC ABW00048;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Human mammaglobin (mgb) protein.  
XX  
KW Cytostatic; epitope; mammaglobin; mgb; therapy; breast cancer; diagnosis;  
XX vaccine; human.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 13..93  
FT /note= "Pro1-9 mammaglobin peptide"  
FT Region 37..75  
FT /note= "Pro-20 mammaglobin peptide"  
FT Region 81..93  
FT /note= "Glob-2 mammaglobin peptide"  
XX  
PN US2002082216-A1.  
XX  
PD 27-JUN-2002.  
XX  
XX 08-JAN-2001; 2001US-00757417.  
XX  
XX 26-MAY-2000; 2000US-00580376.  
XX  
XX (FANG/) FANGER G R.  
XX (FOYT/) FOY T M.  
XX (HOUG/) HOUGHTON R L.  
XX (REED/) REED S G.  
XX  
PI Fanger GR, Foy TM, Houghton RL, Reed SG;  
XX  
DR WPI; 2003-776615/73.  
XX  
XX New polypeptide epitopes of human mammaglobin useful in inhibiting  
PT development of breast cancer and in breast cancer diagnosis and  
PT monitoring, and to produce antibodies also useful in breast cancer  
PT therapy, diagnosis and monitoring.  
XX  
PS Example 1; Fig 2; 66pp; English.  
XX  
XX The present invention provides polypeptide epitopes of human mammaglobin  
CC (mgb) useful for the therapy, diagnosis and monitoring of breast cancer.  
CC The invention is related to specific epitopes of mammaglobin, to  
CC antibodies and immune cells that recognise such epitopes and to methods  
CC for detecting mammaglobin in patient serum. These peptides, antibodies  
CC and cells may be useful in vaccines and pharmaceutical compositions for  
CC prevention and treatment of breast cancer. The invention is also useful  
CC to detect and /or monitor the progression of breast cancer. The present  
CC sequence is human mammaglobin (mgb) protein  
XX  
SQ Sequence 93 AA;

Query Match 100.0%; Score 475; DB 7; Length 93;  
Best Local Similarity 100.0%; Pred. No. 6.1e-47;  
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKLLMVLMLAALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
Db 1 MKLLMVLMLAALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
QY 61 ELKECFLNQTDETLSNVVFQMQLIYDSSLCDLF 93

Db 61 ELKECFLNQTDETLSNVVFQMQLIYDSSLCDLF 93  
RESULT 19  
ADL93165  
ID ADL93165 standard; protein; 93 AA.  
XX  
AC ADL93165;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Human breast cancer-associated polypeptide #22.  
XX  
KW gene therapy; protein therapy; vaccine; breast cancer; cancer; human.  
XX  
OS Homo sapiens.  
XX  
PN US2003166022-A1.  
XX  
PD 04-SEP-2003.  
XX  
PF 15-APR-2002; 2002US-00124805.  
XX  
PR 28-DEC-1998; 98US-00222575.  
PR 02-APR-1999; 99US-00285480.  
PR 23-JUN-1999; 99US-00339338.  
PR 02-SEP-1999; 99US-00389681.  
PR 03-NOV-1999; 99US-00433826.  
PR 17-APR-2000; 2000US-00551621.  
PR 08-JUN-2000; 2000US-00590751.  
PR 22-JUN-2000; 2000US-00604287.  
PR 20-JUL-2000; 2000US-00620405.  
PR 13-APR-2001; 2001US-00834759.  
PR 07-DEC-2001; 2001US-00007805.  
PR 13-FEB-2002; 2002US-00076622.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
XX Houghton RL, Sleath PR, Persing DH;  
XX  
DR WPI; 2003-874918/81.  
XX  
XX An isolated oncogenic polypeptide useful for preventing, diagnosing and  
PT treating breast cancer.  
XX  
PS Disclosure; SEQ ID NO 503; 294pp; English.  
XX  
CC The invention relates to an isolated breast cancer-associated  
CC polypeptide. The polypeptide may be used for the diagnosis and treatment  
CC of breast cancers. The methods are useful for detecting the presence of a  
CC cancer in a patient and treating a cancer in a patient. The present  
CC sequence represents the amino acid sequence of a human breast cancer-  
CC associated polypeptide.  
XX  
SQ Sequence 93 AA;

Query Match 100.0%; Score 475; DB 7; Length 93;  
Best Local Similarity 100.0%; Pred. No. 6.1e-47;  
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKLLMVLMLAALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
Db 1 MKLLMVLMLAALSOHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
QY 61 ELKECFLNQTDETLSNVVFQMQLIYDSSLCDLF 93  
Db 61 ELKECFLNQTDETLSNVVFQMQLIYDSSLCDLF 93

RESULT 20  
ADN04238  
ID ADN04238 standard; protein; 93 AA.

```
XX AC ADN04238;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #314.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT;
XX PI Wu TD;
XX DR WPI; 2004-305105/28.
XX DR N-PSDB; ADN04237.
XX PT New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX PS Claim 9; SEQ ID NO 632; 3069pp; English.
XX CC The invention relates to novel polynucleotide and polypeptides for
XX CC treating psoriasis or a sequence having at least 80% identity to the
XX CC above sequences. The nucleic acid is useful for preparing a composition
XX CC for diagnosing or treating psoriasis in a mammal. This sequence
XX CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 93 AA;

Query Match 100.0%; Score 475; DB 8; Length 93;
Best Local Similarity 100.0%; Pred. No. 6.1e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDDNATTNAID 60
Db 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDDNATTNAID 60

Qy 61 ELKECFNLQTDLTLSNVEVFMQLIYDSSLCDLF 93
Db 61 ELKECFNLQTDLTLSNVEVFMQLIYDSSLCDLF 93

RESULT 21
ADQ17425
ID ADQ17425 standard; protein; 93 AA.
XX AC ADQ17425;
XX DT 26-AUG-2004 (first entry)
XX DE Human soft tissue sarcoma-upregulated protein - SEQ ID 242.
XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX OS Homo sapiens.
XX PN WO2004048938-A2.
XX PD 10-JUN-2004.
XX PR 26-NOV-2003; 2003WO-US038193.

XX PR 26-NOV-2002; 2002US-0429739P.
XX PA (PROT-) PROTEIN DESIGN LABS INC.
XX PI Aziz N, Ginsburg WM, Zlotnik A;
XX DR WPI; 2004-441208/41.
XX PT Early detection of soft tissue sarcoma comprises determining expression
XX PT of a gene in a first soft tissue sample and a normal soft tissue sample
XX PT and comparing the gene expression, also useful in treating soft tissue
XX PT sarcoma.
XX PS Example 2; SEQ ID NO 242; 210pp; English.
XX CC The invention relates to a novel method for detecting soft tissue sarcoma
XX CC which comprises obtaining a first soft tissue sample from an individual
XX CC and a normal soft tissue sample from the same or different individual,
XX CC determining the expression of a gene in both samples and comparing the
XX CC expression of the gene in both soft tissue samples, where a higher level
XX CC of protein expression in the first soft tissue sample indicates the
XX CC presence of soft tissue sarcoma. The method of the invention has
XX CC cytotatic applications and may be useful for detecting soft tissue
XX CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX CC acid sequences may be useful in diagnostic and screening applications.
XX CC The current sequence is that of a human soft tissue sarcoma-upregulated
XX CC protein of the invention. The current sequence is not shown within the
XX CC specification per se but was submitted in CD format by the inventor.
XX SQ Sequence 93 AA;

Query Match 100.0%; Score 475; DB 8; Length 93;
Best Local Similarity 100.0%; Pred. No. 6.1e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDDNATTNAID 60
Db 1 MKLLVLMALALSOHCYAGSGCPLENNVISKTINPQVSKTEYKELLQEFIDDNATTNAID 60

Qy 61 ELKECFNLQTDLTLSNVEVFMQLIYDSSLCDLF 93
Db 61 ELKECFNLQTDLTLSNVEVFMQLIYDSSLCDLF 93

RESULT 22
ABM81260
ID ABM81260 standard; protein; 93 AA.
XX AC ABM81260;
XX DT 18-NOV-2004 (first entry)
XX DE Tumour-associated antigenic target (TAT) polypeptide PRO2018, SEQ.3246.
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX KW tumour; diagnosis; cell proliferative disorder; breast cancer;
XX KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX KW central nervous system cancer; bladder cancer; pancreatic cancer;
XX KW cervical cancer; melanoma; leukaemia; hybridisation probe;
XX KW chromosome identification; chromosome mapping; gene mapping;
XX KW gene therapy; cytostatic.
XX OS Homo sapiens.
XX PN WO2004030615-A2.
XX PD 15-APR-2004.
XX PF 29-SEP-2003; 2003WO-US028547.
XX PR 02-OCT-2002; 2002US-0414971P.
```

PA (GETH ) GENENTECH INC.  
PI Wu TD, Zhang Z, Zhou Y;  
XX  
DR WPI; 2004-347921/32.  
DR N-PSDB; ACN39229.  
XX  
PT New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX  
PS Claim 12; SEQ ID NO 3246; 7273pp; English.  
XX  
CC The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT polypeptide of the invention  
XX  
SQ Sequence 93 AA;  
  
Query Match 100.0%; Score 475; DB 8; Length 93;  
Best Local Similarity 100.0%; Pred. No. 6.1e-47;  
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKLLMVLMLAALSQHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
DB 1 MKLLMVLMLAALSQHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
  
QY 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93  
DB 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93  
  
RESULT 23  
ABG94688  
ID ABG94688 standard; protein; 101 AA.  
XX  
AC ABG94688;  
XX  
DT 02-DEC-2002 (first entry)  
XX  
DE Human mammaglobin/C-terminal His tag.  
XX  
KW Human; epitope; mammaglobin; breast cancer; cytostatic; T cell; CD4+;  
KW CD8+; antigen; His tag.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FN WO200253017-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 08-JAN-2002; 2002WO-US003057.  
XX  
PF 08-JAN-2001; 2001US-00757417.  
PR

PR 08-NOV-2001; 2001US-00008045.  
XX (CORI-) CORIXA CORP.  
XX  
PI Fling SP, Foy TM, Clapper JD, Wang A, Johnson JC, McNeill PD;  
PI Sutherland RA;  
XX  
DR WPI; 2002-706844/76.  
DR N-PSDB; ABS71805.  
XX  
PT Novel polypeptides comprise one or more human mammaglobin epitopes and  
PT polynucleotides encoding the polypeptides, useful for preventing and  
PT treating breast cancers.  
XX  
PS Claim 22; Page 118-119; 121pp; English.  
XX  
CC The invention relates to an isolated polypeptide (I) comprising 7-30  
CC consecutive amino acid residues of human mammaglobin, where one or more  
CC mammaglobin-specific T cells specifically reacts with (I). Also included  
CC are (1) a composition comprising (I), in combination with a  
CC physiologically acceptable carrier or immunostimulant; (2) a diagnostic  
CC kit, comprising (I) and a detection reagent comprising a reporter group;  
CC (3) removing tumour cells from a biological sample, by contacting a  
CC biological sample with T cells that specifically react with (I), under  
CC conditions and for a time sufficient to permit the removal of cells  
CC expressing mammaglobin or a peptide epitope from the sample; (4) an  
CC isolated T cell population (II), comprising T cells prepared using (I);  
CC and (5) a composition (C2) comprising a polynucleotide comprising a  
CC sequence encoding human mammaglobin (or variant or tagged with an  
CC affinity tag), or a polypeptide comprising a human mammaglobin (or  
CC variant or tagged with an affinity tag), in combination with an  
CC immunostimulant. The peptide is useful for inhibiting the development of  
CC breast cancer in a patient, by incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells  
CC isolated from a patient with a mammaglobin epitope, such that T cells  
CC proliferate, administering the proliferated T cells to the patient,  
CC optionally cloning at least one proliferated T cell and administering at  
CC least one cloned cell, and thus inhibiting the development of breast  
CC cancer in the patient. The compositions and methods are useful for  
CC inhibiting the development of breast cancer in a patient. The present  
CC sequence is a human mammaglobin variant protein or fusion protein  
CC containing an N or C His tag  
XX  
SQ Sequence 101 AA;  
  
Query Match 100.0%; Score 475; DB 5; Length 101;  
Best Local Similarity 100.0%; Pred. No. 6.8e-47;  
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKLLMVLMLAALSQHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
DB 1 MKLLMVLMLAALSQHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNATTNAID 60  
  
QY 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93  
DB 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93  
  
RESULT 24  
ABG94689  
ID ABG94689 standard; protein; 102 AA.  
XX  
AC ABG94689;  
XX  
DT 02-DEC-2002 (first entry)  
XX  
DE Human mammaglobin/N-terminal His tag.  
XX  
KW Human; epitope; mammaglobin; breast cancer; cytostatic; T cell; CD4+;  
KW CD8+; antigen; His tag.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX



PN WO200253017-A2.  
 XX 11-JUL-2002.  
 PD 08-JAN-2002; 2002WO-US003057.  
 PF 08-JAN-2001; 2001US-00757417.  
 PR 08-NOV-2001; 2001US-00008045.  
 XX (CORI-) CORIXA CORP.  
 XX Fling SP, Foy TM, Clapper JD, Wang A, Johnson JC, McNeill PD;  
 PI Sutherland RA;  
 XX WPI; 2002-706844/76.  
 DR N-PSDB; ABS71804.  
 XX Novel polypeptides comprise one or more human mamaglobin epitopes and  
 PT polynucleotides encoding the polypeptides, useful for preventing and  
 PT treating breast cancers.  
 XX Claim 22; Page 119; 121pp; English.  
 PS The invention relates to an isolated polypeptide (I) comprising 7-30  
 CC consecutive amino acid residues of human mamaglobin, where one or more  
 CC mamaglobin-specific T cells specifically reacts with (I). Also included  
 CC are (1) a composition comprising (I), in combination with a  
 CC physiologically acceptable carrier or immunostimulant; (2) a diagnostic  
 CC kit, comprising (I) and a detection reagent comprising a reporter group;  
 CC (3) removing tumour cells from a biological sample, by contacting a  
 CC biological sample with T cells that specifically react with (I), under  
 CC conditions and for a time sufficient to permit the removal of cells  
 CC expressing mamaglobin or a peptide epitope from the sample; (4) an  
 CC isolated T cell population (II), comprising a polynucleotide comprising a  
 CC sequence encoding human mamaglobin (or variant or tagged with an  
 CC affinity tag), or a polypeptide comprising a human mamaglobin (or  
 CC variant or tagged with an affinity tag), in combination with an  
 CC immunostimulant. The peptide is useful for inhibiting the development of  
 CC breast cancer in a patient, by incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells  
 CC isolated from a patient with a mamaglobin epitope, such that T cells  
 CC proliferate, administering the proliferated T cells to the patient,  
 CC optionally cloning at least one proliferated T cell and administering at  
 CC least one cloned cell, and thus inhibiting the development of breast  
 CC cancer in the patient. The compositions and methods are useful for  
 CC inhibiting the development of breast cancer in a patient. The present  
 CC sequence is a human mamaglobin variant protein or fusion protein  
 CC containing an N or C His tag  
 XX  
 SQ Sequence 102 AA;  
 Query Match 100.0%; Score 475; DB 5; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-47;  
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLLVLMALALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELQEFIDNNATTNAID 60  
 DB 10 MKLLVLMALALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELQEFIDNNATTNAID 69  
 QY 61 ELKECFINQTDETLSNVEVFMQLIYDSSLCDLF 93  
 DB 70 ELKECFINQTDETLSNVEVFMQLIYDSSLCDLF 102  
 RESULT 25  
 ABG94684  
 ID ABG94684 standard; protein; 132 AA.  
 XX ABG94684;  
 AC  
 XX 29-AUG-2003 (revised)  
 DT 02-DEC-2002 (first entry)  
 XX

DE Human mamaglobin/RA12 fusion protein.  
 XX Human; epitope; mamaglobin; breast cancer; cytostatic; T cell; CD4+;  
 KW CD8+; antigen; RA12.  
 XX Homo sapiens.  
 OS Mycobacterium tuberculosis.  
 OS Chimeric.  
 XX WO200253017-A2.  
 XX 11-JUL-2002.  
 PD 08-JAN-2002; 2002WO-US003057.  
 PF 08-JAN-2001; 2001US-00757417.  
 PR 08-NOV-2001; 2001US-00008045.  
 XX (CORI-) CORIXA CORP.  
 XX Fling SP, Foy TM, Clapper JD, Wang A, Johnson JC, McNeill PD;  
 PI Sutherland RA;  
 XX WPI; 2002-706844/76.  
 DR N-PSDB; ABS71801.  
 XX Novel polypeptides comprise one or more human mamaglobin epitopes and  
 PT polynucleotides encoding the polypeptides, useful for preventing and  
 PT treating breast cancers.  
 XX Claim 22; Page 115-116; 121pp; English.  
 PS The invention relates to an isolated polypeptide (I) comprising 7-30  
 CC consecutive amino acid residues of human mamaglobin, where one or more  
 CC mamaglobin-specific T cells specifically reacts with (I). Also included  
 CC are (1) a composition comprising (I), in combination with a  
 CC physiologically acceptable carrier or immunostimulant; (2) a diagnostic  
 CC kit, comprising (I) and a detection reagent comprising a reporter group;  
 CC (3) removing tumour cells from a biological sample, by contacting a  
 CC biological sample with T cells that specifically react with (I), under  
 CC conditions and for a time sufficient to permit the removal of cells  
 CC expressing mamaglobin or a peptide epitope from the sample; (4) an  
 CC isolated T cell population (II), comprising a polynucleotide comprising a  
 CC sequence encoding human mamaglobin (or variant or tagged with an  
 CC affinity tag), or a polypeptide comprising a human mamaglobin (or  
 CC variant or tagged with an affinity tag), in combination with an  
 CC immunostimulant. The peptide is useful for inhibiting the development of  
 CC breast cancer in a patient, by incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells  
 CC isolated from a patient with a mamaglobin epitope, such that T cells  
 CC proliferate, administering the proliferated T cells to the patient,  
 CC optionally cloning at least one proliferated T cell and administering at  
 CC least one cloned cell, and thus inhibiting the development of breast  
 CC cancer in the patient. The compositions and methods are useful for  
 CC inhibiting the development of breast cancer in a patient. The present  
 CC sequence is a human mamaglobin/RA12 fusion protein. (Updated on 29-AUG-  
 CC 2003 to standardise OS field)  
 XX  
 SQ Sequence 132 AA;  
 Query Match 100.0%; Score 475; DB 5; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-47;  
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLLVLMALALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELQEFIDNNATTNAID 60  
 DB 40 MKLLVLMALALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELQEFIDNNATTNAID 99  
 QY 61 ELKECFINQTDETLSNVEVFMQLIYDSSLCDLF 93  
 DB 100 ELKECFINQTDETLSNVEVFMQLIYDSSLCDLF 132

Search completed: September 26, 2005, 08:25:14  
Job time : 111.286 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:18:37 ; Search time 102.857 Seconds  
(without alignments)  
338.415 Million cell updates/sec

Title: US-09-975-502A-8

Perfect score: 458

Sequence: 1 MKFLAVLVLGVSIFLVSQ.....KDIPLVLPKWGDLNGRVCVP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A Geneseq\_16Dec04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	458	100.0	90	2	AAW42088 BS106 pol
2	458	100.0	90	2	AAI59655 Secreted
3	458	100.0	90	2	AAW67900 Human sec
4	458	100.0	90	2	AAI13466 BS106 pol
5	458	100.0	90	3	AAI66753 Membrane-
6	458	100.0	90	3	AB28525 Protein e
7	458	100.0	90	3	AB00184 Breast ca
8	458	100.0	90	4	AAE11221 Human BS1
9	458	100.0	90	4	AAH83832 Amino aci
10	458	100.0	90	4	AAH65988 BS11S pol
11	458	100.0	90	4	AAH53095 Human ang
12	458	100.0	90	4	AAH65276 Human PRO
13	458	100.0	90	5	AAU82641 Human bre
14	458	100.0	90	5	ABH84910 Human pro
15	458	100.0	90	5	ABJ05542 Breast ca
16	458	100.0	90	5	ABH95516 Human ang
17	458	100.0	90	6	ABU58091 Human PRO
18	458	100.0	90	6	ABU59169 Novel hum
19	458	100.0	90	6	ABU82681 Human sec
20	458	100.0	90	6	ABU60600 Human sec
21	458	100.0	90	6	ABU80786 Human PRO
22	458	100.0	90	6	ABO33752 Novel hum
23	458	100.0	90	6	ABU13982 Human PRO
24	458	100.0	90	6	ABU72567 Novel hum
25	458	100.0	90	6	ABR47515 Breast ca

26	458	100.0	90	6	ABU59316 Human sec
27	458	100.0	90	6	ABO26013 Human PRO
28	458	100.0	90	6	ABU82095 Novel hum
29	458	100.0	90	6	ABU59022 Human sec
30	458	100.0	90	6	ABU92400 Novel hum
31	458	100.0	90	6	ABU59465 Novel hum
32	458	100.0	90	6	ABU92231 Novel hum
33	458	100.0	90	6	ABU10937 Human PRO
34	458	100.0	90	6	ABU81689 Novel hum
35	458	100.0	90	6	ABU88628 Human sec
36	458	100.0	90	6	ABO34142 Human PRO
37	458	100.0	90	6	ABJ72275 Human PRO
38	458	100.0	90	6	ADA37905 Human sec
39	458	100.0	90	6	ADA21591 Human sec
40	458	100.0	90	6	ADA10378 Human sec
41	458	100.0	90	6	ADA17922 Human PRO
42	458	100.0	90	6	ADA28030 Human sec
43	458	100.0	90	6	ADA94610 Human sec
44	458	100.0	90	6	ADA43318 Human BS1
45	458	100.0	90	6	ADA38835 Human sec
46	458	100.0	90	6	ABJ72403 Human PRO
47	458	100.0	90	6	ADA92956 Human sec
48	458	100.0	90	6	ABO34298 Human sec
49	458	100.0	90	7	ABO53228 Human sec
50	458	100.0	90	7	ADA22517 Human sec
51	458	100.0	90	7	ABO22598 Human sec
52	458	100.0	90	7	ADA06683 Human sec
53	458	100.0	90	7	ABJ72105 Human mem
54	458	100.0	90	7	ADA39376 Human sec
55	458	100.0	90	7	ADB83586 Novel hum
56	458	100.0	90	7	ADB80692 Novel hum
57	458	100.0	90	7	ADB73233 Novel hum
58	458	100.0	90	7	ADB96402 Human PRO
59	458	100.0	90	7	ADB78315 Novel hum
60	458	100.0	90	7	ADB84963 Human PRO
61	458	100.0	90	7	ADB78069 Novel hum
62	458	100.0	90	7	ADB87135 Human PRO
63	458	100.0	90	7	ADB84717 Human PRO
64	458	100.0	90	7	ADB83832 Novel hum
65	458	100.0	90	7	ADB72987 Novel hum
66	458	100.0	90	7	ADC57874 Human PRO
67	458	100.0	90	7	ADC55238 Human PRO
68	458	100.0	90	7	ADC12105 Human sec
69	458	100.0	90	7	ADC56527 Human PRO
70	458	100.0	90	7	ADC07582 Human sec
71	458	100.0	90	7	ADC11572 Human sec
72	458	100.0	90	7	ADC36825 Human PRO
73	458	100.0	90	7	ADC21815 Human PRO
74	458	100.0	90	7	ADC49846 Novel hum
75	458	100.0	90	7	ADC49045 Novel hum
76	458	100.0	90	7	ADC49562 Novel hum
77	458	100.0	90	7	ADC47423 Novel hum
78	458	100.0	90	7	ADC14694 Novel hum
79	458	100.0	90	7	ADC47168 Novel hum
80	458	100.0	90	7	ADD08226 Novel hum
81	458	100.0	90	7	ADC82051 Human PRO
82	458	100.0	90	7	ADD07693 Novel hum
83	458	100.0	90	7	ADC78043 Novel hum
84	458	100.0	90	7	ADC82584 Human PRO
85	458	100.0	90	7	ADD06278 Novel hum
86	458	100.0	90	7	ADD10477 Human sec
87	458	100.0	90	7	ADD08764 Novel hum
88	458	100.0	90	7	ADC77797 Novel hum
89	458	100.0	90	7	ADD07013 Novel hum
90	458	100.0	90	7	ADD11437 Human sec
91	458	100.0	90	7	ADC83260 Human PRO
92	458	100.0	90	7	ADD50760 Novel hum
93	458	100.0	90	7	ADD51006 Novel hum
94	458	100.0	90	7	ADD5367 Human PRO
95	458	100.0	90	7	ADD37230 Human sec
96	458	100.0	90	7	ADD56325 Human PRO
97	458	100.0	90	7	ADD50487 Human PRO
98	458	100.0	90	7	ADD54763 Human PRO

99 458 100.0 90 7 ADD50241 Human PRO  
100 458 100.0 90 7 ADD51252 Novel hum

## ALIGNMENTS

RESULT 1  
AAW42088  
ID AAW42088 standard; peptide; 90 AA.

AC AAW42088;

XX 25-MAR-2003 (revised)

DT 25-SEP-1998 (first entry)

XX BS106 polypeptide.

DE

XX Human; BS106; breast tissue gene; breast cancer; detection marker.

XX Homo sapiens.

OS

XX WO9818945-A1.

PN

XX 07-MAY-1998.

XX

XX 31-OCT-1997; 97WO-US019836.

XX

XX 31-OCT-1996; 96US-00742067.

XX

XX (ABBO ) ABBOTT LAB.

XX

XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;

PI Granados EN, Hodges SC, Klass MR, Kratochvil JD, Robertsrap J;

PI Russell JC, Stroupe SD;

XX

XX WPI; 1998-272234/24.

XX

XX Detecting BS106 gene products - useful for developing products for

PT detecting, staging, preventing, treating or determining predisposition to

PT e.g. breast cancer.

XX

XX Claim 17; Page 92; 114pp; English.

XX

XX BS106 is a breast tissue gene with which breast cancer and related

CC diseases are associated. The BS106 polypeptides AAW42088-W42092 can be

CC utilized in a variety of of assays for the detection of antibodies to

CC breast tissue. They can also be used as immunogens for the detection of

CC antibodies to breast tissue. The products and methods discussed in this

CC invention can be used for detecting, diagnosing, staging, preventing or

CC treating, or determining predisposition to diseases or conditions of the

CC breast such as breast cancer. (Updated on 25-MAR-2003 to correct PI

CC field.)

XX

SQ Sequence 90 AA;

Query Match

Best Local Similarity 100.0%; Score 458; DB 2; Length 90;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATTAAPTT 60

DB 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATTAAPTT 60

QY 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90

DB 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90

RESULT 2

AAV59655

ID AAV59655 standard; protein; 90 AA.

XX

AC AAY59655;

XX 18-JAN-2000 (first entry)

XX

XX Secreted protein 108-002-5-0-F3-FL.

XX

XX Secreted protein; fingerprint identification technique;

KW chromosome mapping; human; hereditary disease; diagnosis; cancer;

KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;

KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;

KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;

KW hypertension.

XX

OS Homo sapiens.

XX

XX WO9940189-A2.

XX

XX 12-AUG-1999.

XX

XX 09-FEB-1999; 99WO-IB000282.

XX

XX 09-FEB-1998; 98US-0074121P.

XX

XX 13-APR-1998; 98US-0081563P.

XX

XX 10-AUG-1998; 98US-0096116P.

XX

XX 04-SEP-1998; 98US-0099273P.

XX

XX (GEST ) GENSET.

XX

XX Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX

XX WPI; 1999-600966/51.

XX

XX N-PSDB; AA240783.

XX

XX Extended cDNAs useful for expressing secreted proteins and to obtain

PT specific antibodies.

XX

XX Claim 10; Page 190; 244pp; English.

XX

XX This sequence represents a human secreted protein of the invention. The

CC extended cDNAs (or genomic DNAs obtainable from them) may be used to

CC prepare PCR primers and probes. These are useful for forensic matching or

CC positive identification by DNA sequencing. They may also be used in

CC alternative fingerprint identification techniques. Antibodies against the

CC proteins encoded by the extended cDNAs are useful in identification of

CC tissue types or cell species, as well as identifying tissue specific

CC soluble proteins. The sequences can be used for chromosome mapping and

CC identification of genes associated with hereditary diseases or drug

CC response. signal sequences from the cDNAs can be used in construction of

CC secretion vectors. Other sequences derived from the extended cDNAs can be

CC used to clone upstream genomic DNA sequences including promoters. This is

CC in turn useful for identifying proteins that interact with promoter

CC sequences. Some of the proteins may be useful in diagnosing and treating

CC several disorders including, but not limited to: cancer, hyperlipidaemia,

CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and

CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,

CC amino acidurias, hypoglycaemia, male rat infertility and myopathies

XX

SQ Sequence 90 AA;

Query Match

Best Local Similarity 100.0%; Score 458; DB 2; Length 90;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATTAAPTT 60

DB 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATTAAPTT 60

QY 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90

DB 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90

RESULT 3

AAW67900  
 ID AAW67900 standard; protein; 90 AA.  
 XX  
 AC AAW67900;  
 XX  
 DT 25-MAR-1999 (first entry)  
 XX  
 DE Human secreted protein encoded by gene 34 clone HMOJ64.  
 XX  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W09842738-A1.  
 XX  
 PD 01-OCT-1998.  
 XX  
 XX 19-MAR-1998; 98WO-US005311.  
 XX  
 PR 21-MAR-1997; 97US-0041276P.  
 PR 21-MAR-1997; 97US-0041277P.  
 PR 21-MAR-1997; 97US-0041281P.  
 PR 21-MAR-1997; 97US-0042344P.  
 PR 30-MAY-1997; 97US-0048069P.  
 PR 30-MAY-1997; 97US-0048094P.  
 PR 30-MAY-1997; 97US-0048095P.  
 PR 30-MAY-1997; 97US-0048096P.  
 PR 30-MAY-1997; 97US-0048099P.  
 PR 30-MAY-1997; 97US-0048131P.  
 PR 30-MAY-1997; 97US-0048135P.  
 PR 30-MAY-1997; 97US-0048154P.  
 PR 30-MAY-1997; 97US-0048160P.  
 PR 30-MAY-1997; 97US-0048186P.  
 PR 30-MAY-1997; 97US-0048187P.  
 PR 30-MAY-1997; 97US-0048188P.  
 PR 30-MAY-1997; 97US-0048351P.  
 PR 30-MAY-1997; 97US-0048352P.  
 PR 30-MAY-1997; 97US-0048355P.  
 PR 30-MAY-1997; 97US-0050937P.  
 PR 05-AUG-1997; 97US-0054804P.  
 PR 19-AUG-1997; 97US-0056370P.  
 PR 02-OCT-1997; 97US-0060862P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Young P, Greene JM, Ferris AM, Ruben SM, Rosen CA, Duan R, Hu J;  
 XX Florence KA, Olsen HS, Ebner R, Brewer LA, Moore PA, Shi Y;  
 XX Lafleur DW, Ni J;  
 XX  
 XX WPI; 1999-070066/06.  
 XX DR N-PSDB; AAX00704.  
 XX  
 XX New isolated human genes and the secreted polypeptides they encode -  
 XX useful for diagnosis and treatment of e.g. cancers, neurological  
 XX disorders, immune diseases, inflammation or blood disorders.  
 XX  
 XX Claim 11; Page 328-329; 385pp; English.  
 XX  
 XX This sequence represents a secreted human protein encoded by the gene  
 XX clone detailed in the descriptor line. The gene can be used to generate  
 XX fusion proteins by linking to the gene to a human immunoglobulin Fc  
 XX portion (e.g. AAX00602) for increasing the stability of the fused protein  
 XX as compared to the human protein only. The invention relates to 87 novel  
 XX genes and their fragments (nucleic acid sequences: AAX00611-X00724; amino  
 XX acid sequences AAW67807-W68004) which are useful for preventing, treating  
 XX or ameliorating medical conditions e.g. by protein or gene therapy. Also,

CC pathological conditions can be diagnosed by determining the amount of the  
 CC new polypeptides in a sample or by determining the presence of mutations  
 CC in the new polynucleotides. Specific uses are described for each of the  
 CC 87 polynucleotides, based on which tissues they are most highly expressed  
 CC in (see AAX00611 for described uses)  
 XX  
 XX Sequence 90 AA;  
 XX  
 Query Match 100.0%; Score 458; DB 2; Length 90;  
 Best Local Similarity 100.0%; Pred. NO. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTTAAATTTAAPTT 60  
 DB 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTTAAATTTAAPTT 60  
 QY 61 ATTAATAATARKDIPVLPKWVGDLFNGRVCP 90  
 DB 61 ATTAATAATARKDIPVLPKWVGDLFNGRVCP 90  
 RESULT 4  
 AAY13466  
 ID AAY13466 standard; protein; 90 AA.  
 XX  
 AC AAY13466;  
 XX  
 DT 26-JUL-1999 (first entry)  
 XX  
 DE BS106 polypeptide sequence.  
 XX  
 KW BS106; breast; marker; detection; diagnosis; breast cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W09923230-A1.  
 XX  
 PD 14-MAY-1999.  
 XX  
 PF 19-OCT-1998; 98WO-US022020.  
 XX  
 PR 31-OCT-1997; 97US-00962094.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 XX Billings-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;  
 XX Granados E, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp L;  
 XX Russell JC, Stroupe SD;  
 XX WPI; 1999-326991/27.  
 XX DR N-PSDB; AAX55581, AAX55582.  
 XX  
 XX New isolated BS106 nucleic acids.  
 XX  
 XX Claim 17; Page 91; 105pp; English.  
 XX  
 XX The invention provides BS106 polynucleotides (AAX55578-82) and  
 XX polypeptides (AAY13466-470). The BS106 nucleic acids correspond to a  
 XX breast tissue gene and can be used as markers for breast tissue disease.  
 XX The products can be used for detecting, diagnosing, staging, monitoring,  
 XX prognosticating, preventing or treating, or determining diseases or  
 XX conditions associated with BS106, especially breast cancer  
 XX  
 XX Sequence 90 AA;  
 XX  
 Query Match 100.0%; Score 458; DB 2; Length 90;  
 Best Local Similarity 100.0%; Pred. NO. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTTAAATTTAAPTT 60  
 DB 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTTAAATTTAAPTT 60

Oy	61	ATTAATTARKDIPVLPKWVGDLNCRVCP	90	PR	19-JUN-1998;	98US-0089948P.
Db	61	ATTAASTTARKDIPVLPKWVGDLNCRVCP	90	PR	19-JUN-1998;	98US-0089952P.
				PR	22-JUN-1998;	98US-0090246P.
				PR	22-JUN-1998;	98US-0090252P.
				PR	22-JUN-1998;	98US-0090254P.
				PR	23-JUN-1998;	98US-0090349P.
				PR	23-JUN-1998;	98US-0090355P.
				PR	24-JUN-1998;	98US-0090429P.
				PR	24-JUN-1998;	98US-0090431P.
				PR	24-JUN-1998;	98US-0090435P.
				PR	24-JUN-1998;	98US-0090444P.
				PR	24-JUN-1998;	98US-0090445P.
				PR	24-JUN-1998;	98US-0090461P.
				PR	24-JUN-1998;	98US-0090472P.
				PR	24-JUN-1998;	98US-0090535P.
				PR	24-JUN-1998;	98US-0090538P.
				PR	24-JUN-1998;	98US-0090540P.
				PR	24-JUN-1998;	98US-0090557P.
				PR	25-JUN-1998;	98US-0090676P.
				PR	25-JUN-1998;	98US-0090688P.
				PR	25-JUN-1998;	98US-0090690P.
				PR	25-JUN-1998;	98US-0090691P.
				PR	25-JUN-1998;	98US-0090694P.
				PR	25-JUN-1998;	98US-0090695P.
				PR	25-JUN-1998;	98US-0090696P.
				PR	26-JUN-1998;	98US-0090862P.
				PR	26-JUN-1998;	98US-0090863P.
				PR	01-JUL-1998;	98US-0091358P.
				PR	01-JUL-1998;	98US-0091360P.
				PR	02-JUL-1998;	98US-0091478P.
				PR	02-JUL-1998;	98US-0091486P.
				PR	02-JUL-1998;	98US-0091519P.
				PR	02-JUL-1998;	98US-0091544P.
				PR	02-JUL-1998;	98US-0091626P.
				PR	02-JUL-1998;	98US-0091628P.
				PR	02-JUL-1998;	98US-0091633P.
				PR	02-JUL-1998;	98US-0091646P.
				PR	02-JUL-1998;	98US-0091673P.
				PR	07-JUL-1998;	98US-0091978P.
				PR	07-JUL-1998;	98US-0091982P.
				PR	09-JUL-1998;	98US-0092182P.
				PR	10-JUL-1998;	98US-0092472P.
				PR	20-JUL-1998;	98US-0093339P.
				PR	30-JUL-1998;	98US-0094651P.
				PR	04-AUG-1998;	98US-0095282P.
				PR	04-AUG-1998;	98US-0095285P.
				PR	04-AUG-1998;	98US-0095301P.
				PR	04-AUG-1998;	98US-0095302P.
				PR	04-AUG-1998;	98US-0095318P.
				PR	04-AUG-1998;	98US-0095321P.
				PR	04-AUG-1998;	98US-0095325P.
				PR	10-AUG-1998;	98US-0095916P.
				PR	10-AUG-1998;	98US-0095929P.
				PR	11-AUG-1998;	98US-0096012P.
				PR	11-AUG-1998;	98US-0096143P.
				PR	12-AUG-1998;	98US-0096146P.
				PR	12-AUG-1998;	98US-0096329P.
				PR	17-AUG-1998;	98US-0096757P.
				PR	17-AUG-1998;	98US-0096766P.
				PR	17-AUG-1998;	98US-0096773P.
				PR	17-AUG-1998;	98US-0096791P.
				PR	17-AUG-1998;	98US-0096867P.
				PR	17-AUG-1998;	98US-0096891P.
				PR	17-AUG-1998;	98US-0096894P.
				PR	17-AUG-1998;	98US-0096895P.
				PR	17-AUG-1998;	98US-0096897P.
				PR	18-AUG-1998;	98US-0096949P.
				PR	18-AUG-1998;	98US-0096950P.
				PR	18-AUG-1998;	98US-0096959P.
				PR	18-AUG-1998;	98US-0096960P.
				PR	18-AUG-1998;	98US-0097022P.

RESULT 5  
ID AAY66753  
ID AAY66753 standard; protein; 90 AA.  
AC AAY66753;  
XX  
DT 05-APR-2000 (first entry)  
XX  
XX Membrane-bound protein PRO1160.  
XX  
XX Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;  
KW pharmaceutical; receptor immunoadhesin; gene mapping.  
XX  
XX Homo sapiens.  
XX  
XX WO9963088-A2.  
XX  
PD 09-DEC-1999.  
XX  
XX 02-JUN-1999; 99WO-US012252.  
XX  
XX 02-JUN-1998; 98US-0087607P.  
PR 02-JUN-1998; 98US-0087609P.  
PR 02-JUN-1998; 98US-0087759P.  
PR 03-JUN-1998; 98US-0087827P.  
PR 04-JUN-1998; 98US-0088021P.  
PR 04-JUN-1998; 98US-0088025P.  
PR 04-JUN-1998; 98US-0088028P.  
PR 04-JUN-1998; 98US-0088029P.  
PR 04-JUN-1998; 98US-0088030P.  
PR 04-JUN-1998; 98US-0088033P.  
PR 04-JUN-1998; 98US-0088326P.  
PR 05-JUN-1998; 98US-0088167P.  
PR 05-JUN-1998; 98US-0088202P.  
PR 05-JUN-1998; 98US-0088212P.  
PR 05-JUN-1998; 98US-0088217P.  
PR 09-JUN-1998; 98US-0088655P.  
PR 10-JUN-1998; 98US-0088722P.  
PR 10-JUN-1998; 98US-0088730P.  
PR 10-JUN-1998; 98US-0088734P.  
PR 10-JUN-1998; 98US-0088738P.  
PR 10-JUN-1998; 98US-0088740P.  
PR 10-JUN-1998; 98US-0088741P.  
PR 10-JUN-1998; 98US-0088742P.  
PR 10-JUN-1998; 98US-0088810P.  
PR 10-JUN-1998; 98US-0088824P.  
PR 10-JUN-1998; 98US-0088825P.  
PR 10-JUN-1998; 98US-0088826P.  
PR 11-JUN-1998; 98US-0088858P.  
PR 11-JUN-1998; 98US-0088861P.  
PR 11-JUN-1998; 98US-0088863P.  
PR 11-JUN-1998; 98US-0088876P.  
PR 12-JUN-1998; 98US-0089090P.  
PR 12-JUN-1998; 98US-0089105P.  
PR 16-JUN-1998; 98US-0089440P.  
PR 16-JUN-1998; 98US-0089512P.  
PR 16-JUN-1998; 98US-0089514P.  
PR 17-JUN-1998; 98US-0089532P.  
PR 17-JUN-1998; 98US-0089538P.  
PR 17-JUN-1998; 98US-0089598P.  
PR 17-JUN-1998; 98US-0089599P.  
PR 17-JUN-1998; 98US-0089600P.  
PR 17-JUN-1998; 98US-0089653P.  
PR 18-JUN-1998; 98US-0089801P.  
PR 18-JUN-1998; 98US-0089907P.  
PR 19-JUN-1998; 98US-0089908P.  
PR 19-JUN-1998; 98US-0089947P.

PR 19-AUG-1998; 98US-0097141P.  
 PR 20-AUG-1998; 98US-0097218P.  
 PR 24-AUG-1998; 98US-0097661P.  
 PR 26-AUG-1998; 98US-0097951P.  
 PR 26-AUG-1998; 98US-0097952P.  
 PR 26-AUG-1998; 98US-0097954P.  
 PR 26-AUG-1998; 98US-0097955P.  
 PR 26-AUG-1998; 98US-0097971P.  
 PR 26-AUG-1998; 98US-0097974P.  
 PR 26-AUG-1998; 98US-0097978P.  
 PR 26-AUG-1998; 98US-0097979P.  
 PR 26-AUG-1998; 98US-0097986P.  
 PR 31-AUG-1998; 98US-0098014P.  
 PR 16-SEP-1998; 98US-0098525P.  
 PR 12-JAN-1999; 98US-0100634P.  
 XX 99US-0115565P.

XX (GETH ) GENENTECH INC.

XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;  
 PI Wood WI, Yuan J;

XX N-PSDB; AAZ65099.  
 DR WPI; 2000-072883/06.

XX Membrane-bound proteins and related nucleotide sequences.

XX Claim 12; Fig 282; 822pp; English.

XX The invention provides membrane-bound PRO polypeptides and  
 CC polynucleotides encoding them. The PRO sequences of the invention were  
 CC identified based on extracellular domain homology screening. The PRO  
 CC sequences have homology with proteins including LDL receptors, TIE  
 CC ligands and various enzymes. The membrane-bound proteins and receptor  
 CC molecules are useful as pharmaceutical and diagnostic agents. Receptor  
 CC immunoadhesins, for instance, can be used as therapeutic agents to block  
 CC receptor-ligand interactions. The membrane-bound proteins can also be  
 CC employed for screening of potential peptide or small molecule inhibitors  
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences  
 CC are useful as hybridization probes, in chromosome and gene mapping and in  
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences will  
 CC also be useful for the preparation of PRO polypeptides, especially by  
 CC recombinant techniques

XX SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 3; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYPATGPADDEAPDAETTAATTAAPT 60  
 Db 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYPATGPADDEAPDAETTAATTAAPT 60

Qy 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90  
 Db 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90

RESULT 6  
 AAB28525  
 ID AAB28525 standard; protein; 90 AA.

XX AAB28525;

XX 07-FEB-2001 (first entry)

XX Protein encoded by human breast tumour cDNA clone B511S.

XX Human; breast tumour antigen; cytostatic; immunotherapy; breast cancer;  
 KW vaccine.

XX Homo sapiens.

XX WO200061756-A2.  
 XX 19-OCT-2000.  
 XX 10-APR-2000; 2000WO-US009688.  
 XX 09-APR-1999; 99US-00288950.  
 XX 02-JUL-1999; 99US-00346327.  
 XX (CORI-) CORIXA CORP.

XX Reed SG, Xu J, Dillon DC;

XX WPI; 2000-638568/61.  
 DR N-PSDB; AAC79470.

XX A novel isolated polypeptide comprising an immunogenic portion of a  
 PT breast cancer protein useful in the detection and treatment of breast  
 PT cancer.

XX Claim 2; Page 90; 95pp; English.

XX The present sequence is encoded by a cDNA sequence which was isolated  
 CC from a breast tumour cDNA library. It is provided in a specification  
 CC relating to compounds for immunotherapy and diagnosis of breast cancer.  
 CC Breast tumour antigens and the polynucleotides that encode them may be  
 CC used in the production of a pharmaceutical composition to be used in the  
 CC treatment of breast cancer. Proliferated T cells and incubated antigen  
 CC presenting cells are also required. The polypeptides and polynucleotides  
 CC may also be used to produce a vaccine

XX SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 3; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYPATGPADDEAPDAETTAATTAAPT 60  
 Db 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYPATGPADDEAPDAETTAATTAAPT 60

Qy 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90  
 Db 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90

RESULT 7  
 AAB00184  
 ID AAB00184 standard; protein; 90 AA.

XX AAB00184;

XX 08-FEB-2001 (first entry)

XX Breast cancer protein BCH1.

XX Breast cancer; diagnosis; prognosis; detection; screening; antibody;  
 KW oestrogen receptor; anti-oestrogen; immune response; lymph node;  
 KW metastases; tumour; BCR3; BCQ8; BCQ5; BCH1; BCN1; BCN5; BCQ2; BCX2;  
 KW BCX3; BCA2; BCR2; BCJ7; BCI3; human.

XX Homo sapiens.

XX WO200055629-A2.

XX 21-SEP-2000.

XX 15-MAR-2000; 2000WO-US006952.

XX 15-MAR-1999; 99US-00268865.

XX 12-NOV-1999; 99US-00439878.

XX 12-NOV-1999; 99US-00440370.

PR 15-NOV-1999; 99US-00440493.  
 PR 16-NOV-1999; 99US-00440676.  
 PR 16-NOV-1999; 99US-00440677.  
 PR 29-NOV-1999; 99US-00450810.  
 PR 02-DEC-1999; 99US-00453137.  
 PR 08-MAR-2000; 2000US-00453137.  
 XX  
 PA (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX Mack D, Gish KC;  
 PI  
 XX WPI; 2000-638216/61.  
 DR N-PSDB; AAA54120, AAA54121.  
 XX  
 XX Screening drug candidates for their ability to modulate breast cancer by  
 PT contacting the drug to a cell expressing an expression profile gene and  
 PT determining modulation of expression of the gene.  
 XX  
 XX Disclosure; Fig 34; 258pp; English.  
 PS  
 XX New methods for screening drug candidates are described which comprise  
 CC adding a drug candidate to a cell that expresses a protein selected from  
 CC BCH1, BCA2, BCJ7, BCN1, BCN5, BCQ2, BCQ5, BCR2, BCX2 and BCY3 or their  
 CC fragments and determining the effect of the drug on the expression of  
 CC those proteins. Antibodies to breast cancer genes (specifically BCH1 or  
 CC its fragment (BCH1p1 or BCH1p2)) are useful for inhibiting and treating  
 CC breast cancer in individuals who are non-responsive to anti-oestrogen and  
 CC positive for oestrogen receptor. Compositions comprising BCH1 or a  
 CC nucleic acid encoding BCH1 are useful for eliciting an immune response in  
 CC an individual. The antibodies are also useful for the diagnosis and  
 CC prognosis of breast cancer and for screening compositions which modulate  
 CC the breast cancer phenotype. The method allows rapid and simple detection  
 CC of lymph node metastases  
 XX  
 XX Sequence 90 AA;  
 SQ

Query Match 100.0%; Score 458; DB 3; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
 DB  
 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
 QY 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90  
 DB 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90

RESULT 8  
 AAEL1221  
 ID AAEL1221 standard; protein; 90 AA.  
 XX  
 AC AAEL1221;  
 DT  
 DT 18-DEC-2001 (first entry)  
 XX  
 XX Human BS106 antigenic epitope #1.  
 XX  
 XX Human; BS106 protein; breast cancer; metastasis; gene therapy; tumour;  
 KW epitope.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX WO200165262-A2.  
 FN  
 XX  
 XX 07-SEP-2001.  
 PD  
 XX  
 XX 28-FEB-2001; 2001WO-US006516.  
 PF  
 XX  
 XX 29-FEB-2000; 2000US-00516444.  
 PR  
 XX

PA (ABBO ) ABBOTT LAB.  
 XX  
 PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;  
 PI Granados EN, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp L;  
 PI Russell JC, Stroupe SD;  
 XX  
 DR WPI; 2001-596773/67.  
 XX  
 PT A gene or its fragment which codes for a BS106 polypeptide, useful for  
 PT the detection of a breast disease such as breast cancer.  
 XX  
 PS Claim 1; Page 153; 158pp; English.  
 XX  
 XX The invention relates to human BS106-specific polypeptides and  
 CC polynucleotides. The BS106 polypeptides and antibodies are useful for  
 CC detecting, diagnosing, staging, monitoring, prognosticating, preventing,  
 CC treating or determining the predisposition of an individual to diseases  
 CC and conditions of the breast such as breast cancer. They are also useful  
 CC in the treatment of tumours or metastases. Polynucleotides of the  
 CC invention are useful in drug screening and gene therapy. The present  
 CC sequence is human BS106 antigenic epitope  
 XX  
 XX Sequence 90 AA;  
 SQ

Query Match 100.0%; Score 458; DB 4; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
 DB  
 1 MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
 QY 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90  
 DB 61 ATTAASTTARKDIPVLPKWVGDLPLNGRVCP 90

RESULT 9  
 AAB83832  
 ID AAB83832 standard; protein; 90 AA.  
 XX  
 AC AAB83832;  
 DT  
 DT 23-JUL-2001 (first entry)  
 XX  
 XX Amino acid sequence of a human breast cancer protein designated BCH1.  
 XX Breast cancer protein; BCH1; breast cancer; chromosome 12; 12q12;  
 KW anti-oestrogen therapy.  
 KW  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..19  
 FT /note= "signal peptide"  
 FT  
 XX WO200135811-A2.  
 PN  
 XX  
 XX 25-MAY-2001.  
 PD  
 XX  
 XX 16-NOV-2000; 2000WO-US031736.  
 PF  
 XX  
 XX 16-NOV-1999; 99US-00440676.  
 PR  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.  
 PA Mack D, Gish KC;  
 XX  
 XX WPI; 2001-343670/36.  
 DR N-PSDB; AAF85500.  
 XX  
 XX BCH1 genes and protein useful for the diagnosis and treatment of breast  
 PT cancer.



XX Disclosure; Fig 3; 73pp; English.

XX The present sequence represents a human breast cancer protein, designated

CC BCH1. BCH1 is upregulated in breast cancer tissue, and is found on

CC chromosome 12, cytoband 12q12. BCH1 can be used as an indicator of breast

CC cancer, for determining non-responsiveness to anti-oestrogen therapy and

CC for treating breast cancer. BCH1 nucleic acid is useful for diagnosing

CC breast cancer, and BCH1 is useful for determining the prognosis of breast

CC cancer and for determining whether an individual with breast cancer will

CC be responsive to anti-oestrogen therapy (where the patient is positive

CC for oestrogen receptor), where high levels indicate poor prognosis and

CC non-responsiveness, respectively. BCH1 is also useful for screening for

CC candidate drugs and bioactive agents. Inhibitors, antibodies and

CC antisense sequences of BCH1 are useful for treating breast cancer.

CC Antibodies to BCH1 are useful for localizing a therapeutic moiety (e.g.

CC cytotoxic agent or radioisotope) to breast cancer tissue, and for

CC treating breast cancer

XX

SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 4; Length 90;

Best Local Similarity 100.0%; Pred. No. 5.8e-38;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLVILGVSTIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATAAPT 60

Db 1 MKFLAVLVILGVSTIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATAAPT 60

Qy 61 ATTAASTTARKDIPVLKVGDLNPGRVCP 90

Db 61 ATTAASTTARKDIPVLKVGDLNPGRVCP 90

RESULT 10

AAG65988

ID AAG65988 standard; protein; 90 AA.

XX

AC AAG65988;

DT 11-FEB-2002 (first entry)

XX

DE B5118 polypeptide sequence.

XX

KW Genetic subtraction; DNA microarray analysis; polymerase chain reaction;

KW cancer; B5118.

XX

OS Homo sapiens.

XX

PN WO200175171-A2.

XX

PD 11-OCT-2001.

XX

PF 02-APR-2001; 2001WO-US010631.

XX

PR 03-APR-2000; 2000US-0194241P.

PR 20-JUL-2000; 2000US-0219862P.

PR 27-JUL-2000; 2000US-0221300P.

PR 18-DEC-2000; 2000US-0256592P.

XX

PA (CORI-) CORIXA CORP.

XX

PI Houghton RL, Dillon DC, Molesch DA, Xu J, Zehentner B, Persing DH;

XX

XX WPI: 2001-626449/72.

DR N-PSDB; AAI67224.

DR

XX

PT Identifying tissue (tumor)-specific polynucleotides overexpressed in

PT tissue of interest as compared to control tissue, for detecting cancer

PT cells in patient, comprises DNA microarray analysis or quantitative

PT polymerase chain reaction.

XX

PS Example; Page 116; 127pp; English.

XX The invention relates to identifying tissue-specific polynucleotides (P)

CC that involves performing a genetic subtraction to identify pool of (P)

CC from tissue of interest (TI), performing DNA microarray analysis to

CC identify first subset of polynucleotides (SP1) at least 2-fold over

CC expressed in TI, and performing quantitative polymerase chain reaction

CC (PCR) analysis on SP1 to identify second subset of (P). The method is

CC useful for determining the presence or absence of a cancer cell in a

CC patient, monitoring the progression of cancer in a patient using a

CC biological sample such as blood, serum, lymph nodes, bone marrow, sputum,

CC urine or a tumour biopsy sample. The methods are useful for determining

CC the presence or absence of or monitoring progression of prostate, breast,

CC colon, ovarian, lung, head and neck, lymphoma, leukemia, melanoma, liver,

CC gastric, kidney, bladder, pancreatic or endometrial cancer. The present

CC sequence represents B5118 polypeptide

XX

SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 4; Length 90;

Best Local Similarity 100.0%; Pred. No. 5.8e-38;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLVILGVSTIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATAAPT 60

Db 1 MKFLAVLVILGVSTIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAATAAPT 60

Qy 61 ATTAASTTARKDIPVLKVGDLNPGRVCP 90

Db 61 ATTAASTTARKDIPVLKVGDLNPGRVCP 90

RESULT 11

AAB53095

ID AAB53095 standard; protein; 90 AA.

XX

AC AAB53095;

DT 28-FEB-2001 (first entry)

XX

DE Human angiogenesis-associated protein PRO1160, SEQ ID NO:160.

XX

KW Human; angiogenesis-associated protein; PRO; endothelial cell growth;

KW cardiac hypertrophy; cardiovascular disorder; endothelial disorder;

KW angiogenic disorder; atherosclerosis; osteoporosis; hypertension;

KW myocardial infarction; diabetic retinopathy; rheumatoid arthritis;

KW Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;

KW Alzheimer's disease; Huntington's disease; stroke; drug screening;

KW gene therapy; transgenic animal.

XX

OS Homo sapiens.

XX

PN WO200053753-A2.

XX

PD 14-SEP-2000.

XX

PF 05-JAN-2000; 2000WO-US000219.

XX

PR 08-MAR-1999; 99WO-US0005028.

PR 12-MAR-1999; 99US-0123957P.

PR 14-MAY-1999; 99US-0134287P.

PR 02-JUN-1999; 99WO-US012252.

PR 23-JUN-1999; 99US-0141037P.

PR 20-JUL-1999; 99US-0144758P.

PR 26-JUL-1999; 99US-0145698P.

PR 01-SEP-1999; 99WO-US020111.

PR 08-SEP-1999; 99WO-US020594.

PR 15-SEP-1999; 99WO-US021090.

PR 15-SEP-1999; 99WO-US021547.

PR 05-OCT-1999; 99WO-US023089.

PR 30-NOV-1999; 99WO-US028313.

PR 30-NOV-1999; 99WO-US028409.

PR 02-DEC-1999; 99WO-US028564.

PR 02-DEC-1999; 99WO-US028565.

XX (GETH ) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Kuo SS, Mark MR, Marsters SA, Paoni NF, Pitti RW, Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-090793/10.

DR N-PSDB; AAC97492.

XX New isolated nucleic acid for producing a PRO polypeptide, analyzing genetic disorders and treating cardiovascular, endothelial or angiogenic disorders, such as atherosclerosis, wounds or cancer.

XX Claim 69; Fig 64; 293pp; English.

XX The invention relates to novel human angiogenesis-associated proteins designated PRO proteins (AAB53064-B53097), and to nucleic acids encoding PRO proteins. The invention also relates to vectors and host cells comprising a PRO nucleic acid, the recombinant production of a PRO protein, PRO antibodies specific for a PRO protein, fusion proteins, and compounds which inhibit the expression of a PRO protein, and additionally encompasses methods of identifying modulators of PRO expression or activity; diagnosing a cardiovascular, endothelial or angiogenic disorder, or a susceptibility to such a disorder by detecting mutations in a PRO gene, or the expression level of a PRO gene within a particular tissue; treating a cardiovascular, endothelial or angiogenic disorder via the administration of a PRO protein, PRO nucleic acid, or PRO agonist or antagonist; a retroviral gene therapy vector comprising a PRO nucleic acid; and methods of inhibiting or stimulating endothelial cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the administration of a PRO protein, or an agonist or antagonist thereof. PRO nucleic acids, PRO proteins, antibodies against PRO proteins, PRO agonists and PRO antagonists may be used as therapeutic agents to treat cardiovascular, endothelial or angiogenic disorders, such as atherosclerosis, osteoporosis, myocardial infarction, hypertension, diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis, endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's disease, or stroke. PRO nucleic acids are additionally useful in the recombinant production of PRO proteins, as hybridisation probes to screen libraries to isolate cDNAs with sequence identity to PRO proteins, to map genes encoding PRO proteins, to analyse genetic disorders, and in gene therapy. PRO nucleic acids can also be used to produce transgenic animals useful for the development and screening of potential therapeutic agents. The present sequence represents a PRO protein of the invention

XX Sequence 90 AA;

Query Match 100.0%; Score 458; DB 4; Length 90;

Best Local Similarity 100.0%; Pred. No. 5.8e-38;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLAVLVLLGVSIPLVSAQNPTTAPADTYPATGPAADDEAPDAETTAATTAAPTT 60

Db 1 MKFLAVLVLLGVSIPLVSAQNPTTAPADTYPATGPAADDEAPDAETTAATTAAPTT 60

QY 61 ATTAASTTARKDIPVLPKWGDLPNGRVCP 90

Db 61 ATTAASTTARKDIPVLPKWGDLPNGRVCP 90

RESULT 12

AAB65276

ID AAB65276 standard; protein; 90 AA.

AC AAB65276;

XX 02-APR-2001 (first entry)

XX Human PRO1160 (UNQ590) protein sequence SEQ ID NO:394.

XX Human; secreted and transmembrane protein; PRO; cytostatic; cell death;

KW cancer; chromosomal mapping; gene mapping; tissue typing; diagnostic assay.

KW Homo sapiens.

OS WO200073454-A1.

XX 07-DEC-2000.

XX 30-MAR-2000; 2000WO-US008439.

XX 02-JUN-1999; 99WO-US012252.

PR 23-JUN-1999; 99US-0141037P.

PR 07-JUL-1999; 99US-0143048P.

PR 20-JUL-1999; 99US-0144758P.

PR 26-JUL-1999; 99US-0145698P.

PR 28-JUL-1999; 99US-0146222P.

PR 17-AUG-1999; 99US-0149396P.

PR 15-SEP-1999; 99WO-US021090.

PR 15-SEP-1999; 99WO-US021547.

PR 08-OCT-1999; 99US-0158663P.

PR 30-NOV-1999; 99WO-US028313.

PR 01-DEC-1999; 99WO-US028301.

PR 16-DEC-1999; 99WO-US030095.

PR 20-DEC-1999; 99WO-US030911.

PR 05-JAN-2000; 2000WO-US000219.

PR 06-JAN-2000; 2000WO-US000376.

PR 11-FEB-2000; 2000WO-US000365.

PR 18-FEB-2000; 2000WO-US004341.

PR 22-FEB-2000; 2000WO-US004414.

PR 24-FEB-2000; 2000WO-US004914.

PR 24-FEB-2000; 2000WO-US005004.

PR 02-MAR-2000; 2000WO-US005841.

PR 15-MAR-2000; 2000WO-US006884.

PR 20-MAR-2000; 2000WO-US007377.

XX (GETH ) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL; Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ; Grimaldi CJ, Gurney AL, Kijavini IJ, Napier MA, Pan J, Paoni NF; Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI; Zhang Z;

XX WPI; 2001-032160/04.

DR N-PSDB; AAF44245.

XX PRO polynucleotides used to produce polypeptides used to target bioactive molecules such as toxins, radiolabels or antibodies, to specific cells, to cause targeted cell death.

XX Claim 12; Fig 282; 935pp; English.

XX The present invention describes human secreted and transmembrane PRO proteins. The PRO proteins have cytostatic activity. The PRO proteins can be used for targeted delivery of bioactive molecules, such as toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide sequences, and their fragments, can be used as hybridisation probes, in chromosomal and gene mapping, and in the generation of anti-sense RNA and DNA. They may also be used to produce transgenic animals which are used to develop and screen therapeutically useful reagents. The PRO nucleotide and protein sequence can be used for tissue typing and in treating cancer. Anti-PRO antibodies can be used in diagnostic assays. AAF44270 to AAF44470 represent PCR primers and hybridisation probes used in the isolation of human PRO sequences. AAF44087 to AAF44269 and AAB65154 to AAB65300 represent human PRO polynucleotide and protein sequences given in the exemplification of the present invention

XX Sequence 90 AA;

Query Match 100.0%; Score 458; DB 4; Length 90;

Best Local Similarity 100.0%; Pred. No. 5.8e-38;

Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLLGVSLFLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATAATTAAPTT 60  
Db |||||  
Qy 1 MKFLAVLLGVSLFLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATAATTAAPTT 60  
Db |||||

Qy 61 ATTAASTTARKDIPVLKRWGDLNPNRVC 90  
Db |||||

Qy 61 ATTAASTTARKDIPVLKRWGDLNPNRVC 90  
Db |||||

RESULT 13  
AAU82641  
ID AAU82641 standard; peptide; 90 AA.  
AC AAU82641;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Human breast tumour polypeptide clone #1.  
XX  
KW Human; breast tumour polypeptide; breast cancer; cytostatic;  
KW immunostimulant.  
XX  
OS Homo sapiens.  
XX  
PN WO200198339-A2.  
XX  
PD 27-DEC-2001.  
XX  
XX 12-JUN-2001; 2001WO-US019032.  
XX  
XX 22-JUN-2000; 2000US-00602877.  
PR  
PR 12-OCT-2000; 2000US-00687507.  
PR  
PR 06-FEB-2001; 2001US-00778381.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Reed SG, Xu J, Dillon DC, Retter MW, Harlocker SL;  
XX WPI; 2002-147792/19.  
XX N-PSDB; ABK28975, ABK29014.  
XX  
XX Polynucleotides encoding breast tumor polypeptides, useful for treating  
PT breast cancer or stimulating an immune response.  
PS  
PS Claim 2; Page 142; 150pp; English.  
XX  
XX The invention relates to polynucleotides encoding breast tumour  
CC polypeptides. The sequences are useful for treating cancer, preferably  
CC breast cancer, in a patient or for stimulating an immune response. The  
CC polynucleotides and polypeptides are also useful in the diagnosis and  
CC monitoring of breast cancer. A method for detecting the presence of a  
CC cancer in a patient, comprises obtaining a biological sample from the  
CC patient, contacting the biological sample with a binding agent that binds  
CC to a breast tumour polypeptide, detecting in the sample an amount of  
CC polypeptide that binds to the binding agent, and comparing the amount of  
CC polypeptide to a predetermined cut-off value, therefore determining the  
CC presence of a cancer in the patient. Sequences AAU82641-AAU82655  
CC represent human breast tumour polypeptides of the invention  
XX  
SQ Sequence 90 AA;  
Query Match 100.0%; Score 458; DB 5; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLLGVSLFLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATAATTAAPTT 60  
Db |||||  
Qy 1 MKFLAVLLGVSLFLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATAATTAAPTT 60  
Db |||||

Qy 61 ATTAASTTARKDIPVLKRWGDLNPNRVC 90  
Db |||||

Qy 61 ATTAASTTARKDIPVLKRWGDLNPNRVC 90  
Db |||||

RESULT 14  
ABB84910  
ID ABB84910 standard; protein; 90 AA.  
AC ABB84910;  
XX  
DT 16-MAY-2002 (first entry)  
XX  
XX Human PRO1160 protein sequence SEQ ID NO:188.  
XX  
XX Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;  
KW vulnery; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;  
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;  
KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;  
KW age-related macular degeneration; arterial restenosis; angina;  
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;  
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;  
KW wound healing; chromosome mapping; gene mapping.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200200690-A2.  
PN  
XX  
XX 03-JAN-2002.  
PD  
XX  
XX 20-JUN-2001; 2001WO-US019692.  
PF  
XX  
XX 23-JUN-2000; 2000US-0213637P.  
PR  
XX 20-JUL-2000; 2000US-0219556P.  
PR  
XX 25-JUL-2000; 2000US-0220624P.  
PR  
XX 25-JUL-2000; 2000US-0220664P.  
PR  
XX 02-AUG-2000; 2000WO-US020710.  
PR  
XX 17-AUG-2000; 2000US-0222695P.  
PR  
XX 17-AUG-2000; 2000US-00643657.  
PR  
XX 23-AUG-2000; 2000WO-US023522.  
PR  
XX 24-AUG-2000; 2000WO-US023328.  
PR  
XX 07-SEP-2000; 2000US-0230978P.  
PR  
XX 18-SEP-2000; 2000US-00664610.  
PR  
XX 24-OCT-2000; 2000US-00665350.  
PR  
XX 08-NOV-2000; 2000US-0242922P.  
PR  
XX 08-NOV-2000; 2000US-00709238.  
PR  
XX 10-NOV-2000; 2000WO-US030952.  
PR  
XX 01-DEC-2000; 2000WO-US032678.  
PR  
XX 20-DEC-2000; 2000US-00747259.  
PR  
XX 22-DEC-2000; 2000WO-US034956.  
PR  
XX 28-FEB-2001; 2001US-00796498.  
PR  
XX 28-FEB-2001; 2001WO-US006520.  
PR  
XX 01-MAR-2001; 2001WO-US006666.  
PR  
XX 09-MAR-2001; 2001US-00802706.  
PR  
XX 14-MAR-2001; 2001US-00808689.  
PR  
XX 22-MAR-2001; 2001US-00816744.  
PR  
XX 05-APR-2001; 2001US-00828366.  
PR  
XX 10-MAY-2001; 2001US-00854208.  
PR  
XX 10-MAY-2001; 2001US-00854280.  
PR  
XX 25-MAY-2001; 2001US-00866028.  
PR  
XX 25-MAY-2001; 2001US-00866034.  
PR  
XX 25-MAY-2001; 2001WO-US017092.  
PR  
XX 30-MAY-2001; 2001US-00870574.  
PR  
XX 30-MAY-2001; 2001WO-US017443.  
PR  
XX 01-JUN-2001; 2001WO-US017800.  
PR  
XX (GETH ) GENENTECH INC.  
XX  
XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;  
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;  
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;  
XX WPI; 2002-090516/12.  
XX N-PSDB; ABL88165.  
DR

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,  
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial  
PT infarction), endothelial or angiogenic disorders in a mammal.  
XX  
XX Claim 11; Fig 188; 565pp; English.  
XX  
XX ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to  
CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytostatic,  
CC antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic  
CC activities, and can be used in gene therapy. The PRO polynucleotides,  
CC proteins, agonists and antagonists are useful for treating or diagnosing  
CC a cardiovascular, endothelial or angiogenic disorder in a mammal, e.g.  
CC cardiac hypertrophy, trauma, cancer, age-related macular degeneration,  
CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,  
CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour  
CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound  
CC healing. The PRO polynucleotides have applications in molecular biology,  
CC including use as hybridisation probes, and in chromosome and gene  
CC mapping. ABL88259 to ABL88267 represent primers and probes used in the  
CC exemplification of the present invention  
XX  
SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 5; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5,8e-38;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKFLAVLLGVSI FLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
DB 1 MKFLAVLLGVSI FLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
QY 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90  
DB 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90

RESULT 15  
ABJ05542  
ID ABJ05542 standard; protein; 90 AA.  
XX  
AC ABJ05542;  
XX  
DT 14-NOV-2002 (first entry)  
XX  
DE Breast cancer-associated protein 7.  
XX  
KW Breast cancer; breast cancer-associated gene sequence; drug development;  
KW pharmacogenetics; biosensor development.  
XX  
OS Unidentified.  
XX  
PN WO200259377-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 24-JAN-2002; 2002WO-US002242.  
XX  
PR 24-JAN-2001; 2001US-0263965P.  
PR 02-FEB-2001; 2001US-0265928P.  
PR 09-APR-2001; 2001US-00829472.  
PR 04-APR-2001; 2001US-0282698P.  
PR 04-MAY-2001; 2001US-0288590P.  
PR 29-MAY-2001; 2001US-0294443P.  
XX  
XX (BOSB-) EOS BIOTECHNOLOGY INC.  
XX  
XX Mack DH, Gish KC, Afar D;  
XX  
XX WPI; 2002-583738/62.  
DR N-PSDB; ABT07699.  
DR  
XX  
XX Detecting a breast cancer-associated transcript in a patient's cell,

PT useful for diagnosing breast cancer, comprises contacting a biological  
PT sample with a polynucleotide that selectively hybridizes with breast  
PT cancer nucleic acids.  
XX  
XX Disclosure; Page 353; 414pp; English.  
XX  
XX The invention comprises a method of detecting a breast cancer-associated  
CC transcript in a cell from a patient. The method of the invention involves  
CC contacting a biological sample from the patient with a nucleotide that  
CC hybridises to one of the 69 breast cancer-associated gene sequences shown  
CC in the specification. The method of the invention is useful in the  
CC diagnosis or prognosis of breast cancer, and for detecting genes that are  
CC up or down-regulated in breast cancer cells. Genes identified by the  
CC method of the invention can be used in diagnostic purposes and also as  
CC targets for screening for therapeutic compounds that modulate breast  
CC cancer (e.g. hormones or antibodies). Identification of genes that are  
CC over or under expressed in breast cancer can additionally provide high-  
CC resolution, high-sensitivity datasets which can be used in the areas of  
CC diagnostics, therapeutics, drug development, pharmacogenetics, protein  
CC structure and biosensor development. Amino acid sequences ABJ05536 -  
CC ABJ05604 represent the proteins encoded by the 69 breast cancer-  
CC associated genes of the invention  
XX  
SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 5; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5,8e-38;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKFLAVLLGVSI FLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
DB 1 MKFLAVLLGVSI FLVSAQNPTTAAADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
QY 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90  
DB 61 ATTAASTTARKDIPVLPKWGDLNPGRVCP 90

RESULT 16  
AB995516  
ID AB995516 standard; protein; 90 AA.  
XX  
AC AB995516;  
XX  
DT 19-JUL-2002 (first entry)  
XX  
DE Human angiogenesis related protein PRO1160 SEQ ID NO: 188.  
XX  
KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;  
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;  
KW cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary;  
KW antiarteriosclerotic.  
XX  
OS Homo sapiens.  
XX  
PN WO200208284-A2.  
XX  
PD 31-JAN-2002.  
XX  
PF 09-JUL-2001; 2001WO-US021735.  
XX  
PR 20-JUL-2000; 2000US-0219556P.  
PR 25-JUL-2000; 2000US-0220624P.  
PR 25-JUL-2000; 2000US-0220664P.  
PR 28-JUL-2000; 2000WO-US020710.  
PR 02-AUG-2000; 2000US-0222695P.  
PR 17-AUG-2000; 2000WO-US023522.  
PR 23-AUG-2000; 2000WO-US023657.  
PR 24-AUG-2000; 2000WO-US023328.  
PR 07-SEP-2000; 2000US-0230978P.  
PR 18-SEP-2000; 2000US-00664610.  
PR 18-SEP-2000; 2000US-00665350.  
PR 24-OCT-2000; 2000US-0242922P.

PR 08-NOV-2000; 2000US-00709238.  
PR 08-NOV-2000; 2000WO-US030952.  
PR 10-NOV-2000; 2000WO-US030873.  
PR 01-DEC-2000; 2000WO-US032678.  
PR 20-DEC-2000; 2000US-00747259.  
PR 20-DEC-2000; 2000WO-US034956.  
PR 22-JAN-2001; 2001US-00767609.  
PR 28-FEB-2001; 2001US-00796498.  
PR 28-FEB-2001; 2001WO-US006520.  
PR 01-MAR-2001; 2001WO-US006666.  
PR 09-MAR-2001; 2001US-00802706.  
PR 14-MAR-2001; 2001US-00808689.  
PR 22-MAR-2001; 2001US-00816744.  
PR 05-APR-2001; 2001US-00828366.  
PR 10-MAY-2001; 2001US-00854208.  
PR 10-MAY-2001; 2001US-00854280.  
PR 25-MAY-2001; 2001US-00866028.  
PR 25-MAY-2001; 2001US-00866034.  
PR 25-MAY-2001; 2001WO-US017092.  
PR 30-MAY-2001; 2001US-00870574.  
PR 30-MAY-2001; 2001WO-US017443.  
PR 01-JUN-2001; 2001WO-US017800.  
PR 20-JUN-2001; 2001WO-US019692.

XX (GETH ) GENENTECH INC.  
PA (BAKE/) BAKER K P.  
PA (FERR/) FERRARA N.  
PA (GERB/) GERBER H.  
PA (GERR/) GERRITSEN M E.  
PA (GODD/) GODDARD A.  
PA (GODO/) GODOWSKI P J.  
PA (GURN/) GURNEY A L.  
PA (HILL/) HILLAN K J.  
PA (MARS/) MARSTERS S A.  
PA (PANJ/) PAN J.  
PA (PAON/) PAONI N P.  
PA (STEP/) STEPHAN J F.  
PA (WATA/) WATANABE C K.  
PA (WILL/) WILLIAMS P M.  
PA (WOOD/) WOOD W I.

XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF; Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;

XX WPI; 2002-171999/22.  
DR N-PSDB; ABL95654.

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides, useful in diagnosis and treatment of cardiovascular (e.g. myocardial infarction), endothelial or angiogenic disorders in a mammal.

XX Claim 11; Fig 188; 567pp; English.

XX The present invention provides the protein and coding sequences of human PRO proteins. These are useful for treating or diagnosing a cardiovascular, endothelial or angiogenic disorder, including cardiac hypertrophy, trauma, cancer, age-related macular degeneration, atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour angiogenesis (such as breast carcinoma and liver carcinoma) and wound healing. The present sequence is a PRO protein of the invention

XX Sequence 90 AA;

Query Match 100.0%; Score 458; DB 5; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MKFLAVLLGVSIFLVSAQNPTTAAADTYPATGPDDEADPAETTTAAATTAAPTT 60  
Db 1 MKFLAVLLGVSIFLVSAQNPTTAAADTYPATGPDDEADPAETTTAAATTAAPTT 60

Oy 61 ATTAASTTARKDIPVLPKWVGDLPNGRVCP 90  
Db 61 ATTAASTTARKDIPVLPKWVGDLPNGRVCP 90

RESULT 17  
ABUS8091

ID ABUS8091 standard; protein; 90 AA.

XX AC ABUS8091;

XX DT 14-APR-2003 (first entry)

XX DE Human PRO polypeptide #123.

XX KW Human; PRO; cytostatic; tumour; cancer; breast; lung; stomach; liver;  
XX KW horse; cow; dog; cat; sheep; pig; goat; rabbit; ADEPT;  
XX KW antibody-dependent enzyme mediated prodrug therapy.

XX OS Homo sapiens.

XX PN US2003027163-A1.

XX PD 06-FEB-2003.

XX PF 15-NOV-2001; 2001US-00997666.

XX PR 16-JUN-1997; 97US-0049787P.

XX PR 17-OCT-1997; 97US-0062250P.

XX PR 05-NOV-1997; 97WO-US020069.

XX PR 12-NOV-1997; 97US-0065186P.

XX PR 13-NOV-1997; 97US-0065311P.

XX PR 24-NOV-1997; 97US-0066770P.

XX PR 25-FEB-1998; 98US-0075945P.

XX PR 20-MAR-1998; 98US-0078910P.

XX PR 28-APR-1998; 98US-0083322P.

XX PR 07-MAY-1998; 98US-0084600P.

XX PR 28-MAY-1998; 98US-0087106P.

XX PR 02-JUN-1998; 98US-0087607P.

XX PR 02-JUN-1998; 98US-0087609P.

XX PR 02-JUN-1998; 98US-0087759P.

XX PR 03-JUN-1998; 98US-0087827P.

XX PR 04-JUN-1998; 98US-0088021P.

XX PR 04-JUN-1998; 98US-0088025P.

XX PR 04-JUN-1998; 98US-0088026P.

XX PR 04-JUN-1998; 98US-0088028P.

XX PR 04-JUN-1998; 98US-0088030P.

XX PR 04-JUN-1998; 98US-0088033P.

XX PR 04-JUN-1998; 98US-0088326P.

XX PR 05-JUN-1998; 98US-0088167P.

XX PR 05-JUN-1998; 98US-0088202P.

XX PR 05-JUN-1998; 98US-0088212P.

XX PR 05-JUN-1998; 98US-0088217P.

XX PR 09-JUN-1998; 98US-0088655P.

XX PR 10-JUN-1998; 98US-0088734P.

XX PR 10-JUN-1998; 98US-0088738P.

XX PR 10-JUN-1998; 98US-0088742P.

XX PR 10-JUN-1998; 98US-0088810P.

XX PR 10-JUN-1998; 98US-0088824P.

XX PR 10-JUN-1998; 98US-0088826P.

XX PR 11-JUN-1998; 98US-0088858P.

XX PR 11-JUN-1998; 98US-0088861P.

XX PR 11-JUN-1998; 98US-0088876P.

XX PR 12-JUN-1998; 98US-0089105P.

XX PR 16-JUN-1998; 98US-0089440P.

XX PR 16-JUN-1998; 98US-0089512P.

XX PR 16-JUN-1998; 98US-0089514P.

XX PR 17-JUN-1998; 98US-0089532P.

XX PR 17-JUN-1998; 98US-0089538P.

XX PR 17-JUN-1998; 98US-0089598P.

XX PR 17-JUN-1998; 98US-0089599P.

XX PR 17-JUN-1998; 98US-0089600P.

```
PR 17-JUN-1998; 98US-0089853P.
PR 18-JUN-1998; 98US-0089801P.
PR 18-JUN-1998; 98US-0089907P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089947P.
PR 19-JUN-1998; 98US-0089948P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 23-JUN-1998; 98US-0090349P.
PR 23-JUN-1998; 98US-0090355P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090431P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090445P.
PR 24-JUN-1998; 98US-0090472P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 24-JUN-1998; 98US-0090542P.
PR 24-JUN-1998; 98US-0090557P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 01-JUL-1998; 98US-0091360P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091519P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091633P.
PR 02-JUL-1998; 98US-0091646P.
PR 02-JUL-1998; 98US-0091673P.
PR 07-JUL-1998; 98US-0091978P.
PR 07-JUL-1998; 98US-0091982P.
PR 09-JUL-1998; 98US-0092182P.
PR 10-JUL-1998; 98US-0092472P.
PR 20-JUL-1998; 98US-0093339P.
PR 30-JUL-1998; 98US-0094651P.
PR 04-AUG-1998; 98US-0095282P.
PR 04-AUG-1998; 98US-0095285P.
PR 04-AUG-1998; 98US-0095301P.
PR 04-AUG-1998; 98US-0095302P.
PR 04-AUG-1998; 98US-0095318P.
PR 04-AUG-1998; 98US-0095321P.
PR 10-AUG-1998; 98US-0095325P.
PR 10-AUG-1998; 98US-0095816P.
PR 10-AUG-1998; 98US-0095929P.
PR 10-AUG-1998; 98US-0096012P.
PR 11-AUG-1998; 98US-0096143P.
PR 11-AUG-1998; 98US-0096146P.
PR 12-AUG-1998; 98US-0096329P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096768P.
PR 17-AUG-1998; 98US-0096773P.
PR 17-AUG-1998; 98US-0096791P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096894P.
PR 17-AUG-1998; 98US-0096895P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096950P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0096960P.
PR 18-AUG-1998; 98US-0097022P.
```

```
PR 19-AUG-1998; 98US-0097141P.
PR 20-AUG-1998; 98US-0097218P.
PR 24-AUG-1998; 98US-0097661P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0097978P.
PR 26-AUG-1998; 98US-0097979P.
PR 26-AUG-1998; 98US-0097986P.
PR 26-AUG-1998; 98US-0098014P.
PR 31-AUG-1998; 98US-0098525P.
PR 16-SEP-1998; 98US-0100634P.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98US-0100858P.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 21-DEC-1998; 98WO-US025108.
PR 22-DEC-1998; 98US-0113296P.
PR 05-JAN-1999; 98WO-US000106.
PR 08-JAN-1999; 98WO-US005028.
PR 12-MAR-1999; 98US-0123957P.
PR 02-JUN-1999; 98WO-US012252.
PR 23-JUN-1999; 98US-0141037P.
PR 07-JUL-1999; 98US-0143048P.
PR 20-JUL-1999; 98US-0144758P.
PR 26-JUL-1999; 98US-0145698P.
PR 28-JUL-1999; 98US-0146222P.
PR 17-AUG-1999; 98US-0149396P.
PR 15-SEP-1999; 98WO-US021090.
PR 15-SEP-1999; 98US-0201547.
PR 08-OCT-1999; 98US-0158663P.
PR 30-NOV-1999; 98WO-US028313.
PR 01-DEC-1999; 98WO-US028301.
PR 01-DEC-1999; 98WO-US028634.
PR 16-DEC-1999; 98WO-US030095.
PR 20-DEC-1999; 98WO-US030911.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 02-MAR-2000; 2000WO-US005841.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 30-MAR-2000; 2000WO-US008439.
PR 15-MAY-2000; 2000WO-US013358.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 23-JUN-2000; 2000US-0213637P.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 07-SEP-2000; 2000US-0230978P.
```

Query Match 100.0%; Score 458; DB 6; Length 90;

Best Local Similarity 100.0%; Pred. No. 5.8e-38; Mismatches 0; Indels 0; Gaps 0;

Matches 90; Conservative 0;

QY 1 MKFLAVLLVLLGVSIFLVSAQNPTTAAADTYPATGPDADDEADTAATTATTAAPTT 60

Db 1 MKFLAVLLVLLGVSIFLVSAQNPTTAAADTYPATGPDADDEADTAATTATTAAPTT 60

QY 61 ATTAASTTARKDIPVLPKWVGDLNGRVCP 90

Db 61 ATTAASTTARKDIPVLPKWVGDLNGRVCP 90

RESULT 18  
 ID ABUS9169  
 AC ABUS9169 standard; protein; 90 AA.  
 AC ABUS9169;  
 XX  
 DT 28-APR-2003 (first entry)  
 XX  
 DE Novel human secreted or transmembrane protein PRO1160.  
 XX  
 KW Human; PRO; hypertrophy of neonatal heart; angiogenesis; wound healing;  
 KW cardiac insufficiency disorder; cancer; tumour; immune response;  
 KW adrenal cortical capillary endothelial growth; c-fos induction;  
 KW vascular endothelial growth factor inhibition; VEGF inhibition;  
 KW endothelial cell growth inhibitor; T-lymphocytes stimulation;  
 KW retinal neurons cell survival; rod photoreceptor cell survival;  
 KW retinal disorder; retinitis pigmentosa; kidney disorder;  
 KW mammalian kidney mesangial cell proliferation; Berger disease;  
 KW dermatitis; herpeticiformis; Crohn's disease; chondrocyte proliferation;  
 KW chondrocyte redifferentiation; sports injury; arthritis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002132252-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 14-NOV-2001; 2001US-00990442.  
 XX  
 PR 16-JUN-1997; 97US-0049787P.  
 PR 17-OCT-1997; 97US-0062250P.  
 PR 05-NOV-1997; 97WO-US020069.  
 PR 12-NOV-1997; 97US-0065186P.  
 PR 13-NOV-1997; 97US-0085311P.  
 PR 24-NOV-1997; 97US-0086770P.  
 PR 25-FEB-1998; 98US-0075945P.  
 PR 20-MAR-1998; 98US-0078910P.  
 PR 28-APR-1998; 98US-0083322P.  
 PR 07-MAY-1998; 98US-0084600P.  
 PR 28-MAY-1998; 98US-0087106P.  
 PR 02-JUN-1998; 98US-0087607P.  
 PR 02-JUN-1998; 98US-0087609P.  
 PR 02-JUN-1998; 98US-0087759P.  
 PR 03-JUN-1998; 98US-0087827P.  
 PR 04-JUN-1998; 98US-0088021P.  
 PR 04-JUN-1998; 98US-0088023P.  
 PR 04-JUN-1998; 98US-0088026P.  
 PR 04-JUN-1998; 98US-0088028P.  
 PR 04-JUN-1998; 98US-0088029P.  
 PR 04-JUN-1998; 98US-0088030P.  
 PR 04-JUN-1998; 98US-0088033P.  
 PR 04-JUN-1998; 98US-0088326P.  
 PR 05-JUN-1998; 98US-0088167P.  
 PR 05-JUN-1998; 98US-0088202P.  
 PR 05-JUN-1998; 98US-0088212P.  
 PR 05-JUN-1998; 98US-0088217P.  
 PR 09-JUN-1998; 98US-0088655P.  
 PR 10-JUN-1998; 98US-0088734P.  
 PR 10-JUN-1998; 98US-0088738P.  
 PR 10-JUN-1998; 98US-0088742P.  
 PR 10-JUN-1998; 98US-0088810P.  
 PR 10-JUN-1998; 98US-0088824P.  
 PR 10-JUN-1998; 98US-0088826P.  
 PR 11-JUN-1998; 98US-0088858P.  
 PR 11-JUN-1998; 98US-0088861P.  
 PR 11-JUN-1998; 98US-0088876P.  
 PR 12-JUN-1998; 98US-0089105P.  
 PR 16-JUN-1998; 98US-0089440P.  
 PR 16-JUN-1998; 98US-0089512P.  
 PR 16-JUN-1998; 98US-0089514P.  
 PR 17-JUN-1998; 98US-0089532P.  
 PR 17-JUN-1998; 98US-0089538P.  
 PR 17-JUN-1998; 98US-0089598P.  
 PR 17-JUN-1998; 98US-0089599P.  
 PR 17-JUN-1998; 98US-0089600P.  
 PR 17-JUN-1998; 98US-0089653P.  
 PR 18-JUN-1998; 98US-0089801P.  
 PR 18-JUN-1998; 98US-0089807P.  
 PR 18-JUN-1998; 98US-0089908P.  
 PR 18-SEP-1998; 98WO-US019330.  
 PR 17-SEP-1998; 98WO-US019437.  
 PR 07-OCT-1998; 98WO-US021141.  
 PR 01-DEC-1998; 98WO-US021108.  
 PR 05-JAN-1999; 99WO-US000106.  
 PR 08-MAR-1999; 99WO-US005028.  
 PR 02-JUN-1999; 99WO-US012252.  
 PR 15-SEP-1999; 99WO-US021090.  
 PR 15-SEP-1999; 99WO-US021547.  
 PR 30-NOV-1999; 99WO-US028313.  
 PR 01-DEC-1999; 99WO-US028301.  
 PR 01-DEC-1999; 99WO-US028634.  
 PR 16-DEC-1999; 99WO-US030095.  
 PR 20-DEC-1999; 99WO-US030911.  
 PR 06-JAN-2000; 2000WO-US000219.  
 PR 06-JAN-2000; 2000WO-US000376.  
 PR 11-FEB-2000; 2000WO-US003565.  
 PR 18-FEB-2000; 2000WO-US004341.  
 PR 22-FEB-2000; 2000WO-US004414.  
 PR 24-FEB-2000; 2000WO-US004914.  
 PR 24-FEB-2000; 2000WO-US005004.  
 PR 02-MAR-2000; 2000WO-US005841.  
 PR 10-MAR-2000; 2000WO-US006319.  
 PR 15-MAR-2000; 2000WO-US006884.  
 PR 20-MAR-2000; 2000WO-US007377.  
 PR 30-MAR-2000; 2000WO-US008439.  
 PR 15-MAY-2000; 2000WO-US013358.  
 PR 17-MAY-2000; 2000WO-US013705.  
 PR 22-MAY-2000; 2000WO-US014042.  
 PR 30-MAY-2000; 2000WO-US014941.  
 PR 02-JUN-2000; 2000WO-US015264.  
 PR 28-JUL-2000; 2000WO-US020710.  
 PR 11-AUG-2000; 2000WO-US022031.  
 PR 23-AUG-2000; 2000WO-US023522.  
 PR 24-AUG-2000; 2000WO-US023328.  
 PR 08-NOV-2000; 2000WO-US030952.  
 PR 01-DEC-2000; 2000WO-US032678.  
 PR 28-FEB-2001; 2001WO-US006520.  
 PR 01-JUN-2001; 2001WO-US017800.  
 PR 20-JUN-2001; 2001WO-US019692.  
 PR 29-JUN-2001; 2001WO-US021066.  
 PR 09-JUL-2001; 2001WO-US021735.  
 PR 28-AUG-2001; 2001US-00941992.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;  
 PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NP;  
 PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;  
 PI Zhang Z;  
 XX  
 DR WPI; 2003-247083/24.  
 DR N-PSDB; ABX80375.  
 XX  
 PT Novel isolated PRO polypeptides e.g., PRO826, PRO1068, PRO1184, PRO1346  
 PT and PRO1375, which stimulate proliferation of stimulated T-lymphocytes  
 PT are therapeutically useful for enhancing immune response and in cancer  
 PT treatments.  
 XX  
 PS Claim 12; Fig 282; 648pp; English.  
 XX  
 CC The invention describes an isolated human PRO polypeptide. The PRO  
 CC polypeptides are useful in detecting PRO polypeptides in a sample, in  
 CC linking a bioactive molecule to a cell expressing a PRO polypeptide, and

CC in modulating at least one biological activity of a cell expressing a PRO  
 CC polypeptide. PRO1312 stimulates hypertrophy of neonatal heart and is thus  
 CC useful for treating cardiac insufficiency disorders. PRO1154 and PRO1186  
 CC stimulate adrenal cortical capillary endothelial growth, and PRO536,  
 CC PRO943, PRO828, PRO826, PRO1068 or PRO535, PRO826, PRO819, PRO1126,  
 CC PRO1360 and PRO1387 induce c-fos in endothelial cells, and are thus  
 CC useful for treating conditions or disorders where angiogenesis would be  
 CC beneficial, e.g. wound healing and antagonist of this polypeptide are  
 CC useful for treating cancerous tumours. PRO812 inhibits vascular  
 CC endothelial growth factor (VEGF) stimulated proliferation of endothelial  
 CC cells and is thus useful for inhibiting endothelial cell growth in  
 CC mammals which would be beneficial in inhibiting tumour growth. PRO826,  
 CC PRO1068, PRO1184, PRO1346 and PRO1375 stimulate proliferation of  
 CC stimulated T-lymphocytes and are therapeutically useful for enhancing  
 CC immune response. PRO828, PRO826, PRO1068 or PRO1132 enhance survival of  
 CC retinal neurons cells (PRO1132 is also enhances survival/proliferation of  
 CC rod photoreceptor cells) and therefore are useful for treating retinal  
 CC disorders of injuries, e.g. retinitis pigmentosa, AMD. PRO819, PRO813  
 CC and PRO1066 induce proliferation of mammalian kidney mesangial cells,  
 CC and therefore are useful for treating kidney disorders associated with  
 CC decreased mesangial cell function such as Berger disease or other  
 CC nephropathies associated with dermatitis, herpeticiformis or Crohn's  
 CC disease. PRO1310, PRO844, PRO1312, PRO1192 and PRO1387 induce the  
 CC proliferation and/or redifferentiation of chondrocytes in culture and are  
 CC thus useful for treating sports injuries, and arthritis. This is the  
 CC amino acid sequence of a novel human PRO protein  
 XX  
 SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 6; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLAVLVGLVSTFLVSAQNPTTAADTYPATGPDADDEAPDAETAAATTAAPTT 60  
 DB 1 MKFLAVLVGLVSTFLVSAQNPTTAADTYPATGPDADDEAPDAETAAATTAAPTT 60  
 QY 61 ATTAASTARKDIPVLFPKWVGDLPNGRVCP 90  
 DB 61 ATTAASTARKDIPVLFPKWVGDLPNGRVCP 90

## RESULT 19

ABU82681  
 ID ABU82681 standard; protein; 90 AA.

AC ABU82681;

XX 26-JUN-2003 (first entry)

DE Human secreted/transmembrane protein PRO1160.

XX Human; PRO; secreted protein; transmembrane protein;  
 KW cardiac insufficiency disorders; angiogenesis; wound healing;  
 KW cancerous tumour; immune response; retinal disorder; sight loss;  
 KW retinitis pigmentosa; age-related macular degeneration; AMD;  
 KW kidney disorder; Berger disease; nephropathy; dermatitis; herpeticiformis;  
 KW Crohn's disease; sports injury; arthritis.

XX Homo sapiens.

OS US2003032023-A1.

PN 13-FEB-2003.

PD 14-NOV-2001; 2001US-00990711.

XX 16-JUN-1997; 97US-0049787P.

PR 17-OCT-1997; 97US-0062250P.

PR 05-NOV-1997; 97WQ-USQ20069.

PR 12-NOV-1997; 97US-0065186P.

PR 13-NOV-1997; 97US-0065311P.

PR 24-NOV-1997; 97US-0066770P.

PR 25-FEB-1998; 98US-0075945P.  
 PR 20-MAR-1998; 98US-0078310P.  
 PR 28-APR-1998; 98US-0083322P.  
 PR 07-MAY-1998; 98US-0084600P.  
 PR 28-MAY-1998; 98US-0087106P.  
 PR 02-JUN-1998; 98US-0087607P.  
 PR 02-JUN-1998; 98US-0087609P.  
 PR 02-JUN-1998; 98US-0087759P.  
 PR 03-JUN-1998; 98US-0087827P.  
 PR 04-JUN-1998; 98US-0088021P.  
 PR 04-JUN-1998; 98US-0088025P.  
 PR 04-JUN-1998; 98US-0088026P.  
 PR 04-JUN-1998; 98US-0088028P.  
 PR 04-JUN-1998; 98US-0088029P.  
 PR 04-JUN-1998; 98US-0088030P.  
 PR 04-JUN-1998; 98US-0088033P.  
 PR 04-JUN-1998; 98US-0088326P.  
 PR 05-JUN-1998; 98US-0088167P.  
 PR 05-JUN-1998; 98US-0088202P.  
 PR 05-JUN-1998; 98US-0088212P.  
 PR 05-JUN-1998; 98US-0088217P.  
 PR 09-JUN-1998; 98US-0088655P.  
 PR 10-JUN-1998; 98US-0088734P.  
 PR 10-JUN-1998; 98US-0088738P.  
 PR 10-JUN-1998; 98US-0088742P.  
 PR 10-JUN-1998; 98US-0088810P.  
 PR 10-JUN-1998; 98US-0088824P.  
 PR 10-JUN-1998; 98US-0088826P.  
 PR 11-JUN-1998; 98US-0088858P.  
 PR 11-JUN-1998; 98US-0088861P.  
 PR 11-JUN-1998; 98US-0088876P.  
 PR 12-JUN-1998; 98US-0089105P.  
 PR 16-JUN-1998; 98US-0089440P.  
 PR 16-JUN-1998; 98US-0089512P.  
 PR 16-JUN-1998; 98US-0089514P.  
 PR 17-JUN-1998; 98US-0089532P.  
 PR 17-JUN-1998; 98US-0089538P.  
 PR 17-JUN-1998; 98US-0089598P.  
 PR 17-JUN-1998; 98US-0089599P.  
 PR 17-JUN-1998; 98US-0089600P.  
 PR 17-JUN-1998; 98US-0089653P.  
 PR 18-JUN-1998; 98US-0089801P.  
 PR 18-JUN-1998; 98US-0089907P.  
 PR 18-JUN-1998; 98US-0089908P.  
 PR 19-JUN-1998; 98US-0089947P.  
 PR 19-JUN-1998; 98US-0089948P.  
 PR 19-JUN-1998; 98US-0089952P.  
 PR 22-JUN-1998; 98US-0090246P.  
 PR 22-JUN-1998; 98US-0090252P.  
 PR 22-JUN-1998; 98US-0090254P.  
 PR 23-JUN-1998; 98US-0090349P.  
 PR 23-JUN-1998; 98US-0090355P.  
 PR 24-JUN-1998; 98US-0090429P.  
 PR 24-JUN-1998; 98US-0090431P.  
 PR 24-JUN-1998; 98US-0090435P.  
 PR 24-JUN-1998; 98US-0090444P.  
 PR 24-JUN-1998; 98US-0090445P.  
 PR 24-JUN-1998; 98US-0090472P.  
 PR 24-JUN-1998; 98US-0090535P.  
 PR 24-JUN-1998; 98US-0090540P.  
 PR 24-JUN-1998; 98US-0090542P.  
 PR 24-JUN-1998; 98US-0090577P.  
 PR 25-JUN-1998; 98US-0090676P.  
 PR 25-JUN-1998; 98US-0090678P.  
 PR 25-JUN-1998; 98US-0090690P.  
 PR 25-JUN-1998; 98US-0090694P.  
 PR 25-JUN-1998; 98US-0090695P.  
 PR 25-JUN-1998; 98US-0090696P.  
 PR 26-JUN-1998; 98US-0090862P.  
 PR 26-JUN-1998; 98US-0090863P.  
 PR 01-JUL-1998; 98US-0091360P.  
 PR 02-JUL-1998; 98US-0091544P.  
 PR 02-JUL-1998; 98US-0091478P.



```
PR 02-JUL-1998; 98US-0091519P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091633P.
PR 02-JUL-1998; 98US-0091646P.
PR 02-JUL-1998; 98US-0091673P.
PR 07-JUL-1998; 98US-0091978P.
PR 07-JUL-1998; 98US-0091982P.
PR 09-JUL-1998; 98US-0092182P.
PR 10-JUL-1998; 98US-0092472P.
PR 20-JUL-1998; 98US-0093339P.
PR 30-JUL-1998; 98US-0094651P.
PR 04-AUG-1998; 98US-0095282P.
PR 04-AUG-1998; 98US-0095285P.
PR 04-AUG-1998; 98US-0095301P.
PR 04-AUG-1998; 98US-0095302P.
PR 04-AUG-1998; 98US-0095318P.
PR 04-AUG-1998; 98US-0095321P.
PR 04-AUG-1998; 98US-0095325P.
PR 10-AUG-1998; 98US-0095916P.
PR 10-AUG-1998; 98US-0095929P.
PR 10-AUG-1998; 98US-0096012P.
PR 11-AUG-1998; 98US-0096143P.
PR 11-AUG-1998; 98US-0096146P.
PR 12-AUG-1998; 98US-0096329P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096768P.
PR 17-AUG-1998; 98US-0096773P.
PR 17-AUG-1998; 98US-0096791P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096894P.
PR 17-AUG-1998; 98US-0096895P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096950P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0096960P.
PR 18-AUG-1998; 98US-0097022P.
PR 19-AUG-1998; 98US-0097141P.
PR 20-AUG-1998; 98US-0097218P.
PR 24-AUG-1998; 98US-0097661P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0097978P.
PR 26-AUG-1998; 98US-0097979P.
PR 26-AUG-1998; 98US-0097986P.
PR 26-AUG-1998; 98US-0098014P.
PR 31-AUG-1998; 98US-0098525P.
PR 16-SEP-1998; 98US-0100634P.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 17-SEP-1998; 98WO-US021141.
PR 07-OCT-1998; 98WO-US025108.
PR 21-DEC-1998; 98US-0113296P.
PR 05-JAN-1999; 99WO-US000106.
PR 08-MAR-1999; 99WO-US005028.
PR 12-MAR-1999; 99US-0123957P.
PR 02-JUN-1999; 99WO-US012252.
PR 23-JUN-1999; 99US-0141037P.
PR 07-JUL-1999; 99US-0143048P.
PR 20-JUL-1999; 99US-0144758P.
PR 26-JUL-1999; 99US-0145698P.
PR 28-JUL-1999; 99US-0146222P.
PR 17-AUG-1999; 99US-0149396P.
PR 15-SEP-1999; 99WO-US021090.
PR 15-SEP-1999; 99WO-US021547.
PR 08-OCT-1999; 99US-0158663P.

PR 30-NOV-1999; 99WO-US028313.
PR 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 02-MAR-2000; 2000WO-US005841.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 30-MAR-2000; 2000WO-US008439.
PR 15-MAY-2000; 2000WO-US013358.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 23-JUN-2000; 2000US-0213637P.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.

Query Match 100.0%; Score 458; DB 6; Length 90;
Best Local Similarity 100.0%; Pred. No. 5.8e-38;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLVLGVSIFLYSAQNPTTAAADTYPATGPADDEAPDAETTAATAAATT 60
Db 1 MKFLAVLVLGVSIFLYSAQNPTTAAADTYPATGPADDEAPDAETTAATAAATT 60

Qy 61 ATTAASTTARKDIPVLPKMWGDLNPNRVCVP 90
Db 61 ATTAASTTARKDIPVLPKMWGDLNPNRVCVP 90

RESULT 20
ABU60600
ID ABU60600 standard; protein; 90 AA.
AC ABU60600;
XX
DT 01-MAY-2003 (first entry)
XX
DE Human secreted/transmembrane protein, #159.
XX
KW Human; PRO; secreted; transmembrane; signal peptide; pharmaceutical;
diagnostic; therapeutic; gene therapy.
XX
OS Homo sapiens.
XX
FN US2002160384-A1.
XX
PD 31-OCT-2002.
XX
PF 14-NOV-2001; 2001US-00992598.
XX
PR 16-JUN-1997; 97US-0049787P.
PR 17-OCT-1997; 97US-0062250P.
PR 05-NOV-1997; 97WO-US020069.
PR 12-NOV-1997; 97US-0065186P.
PR 13-NOV-1997; 97US-0065311P.
PR 24-NOV-1997; 97US-0066770P.
PR 25-FEB-1998; 98US-0075945P.
PR 20-MAR-1998; 98US-0078910P.
PR 28-APR-1998; 98US-0083322P.
PR 07-MAY-1998; 98US-0084600P.
PR 28-MAY-1998; 98US-0087106P.
PR 02-JUN-1998; 98US-0087607P.
PR 02-JUN-1998; 98US-0087609P.
```

Human; PRO polypeptide; secreted and transmembrane protein;  
 anti-PRO antibody; diagnostic assay; gene expression; tumour; cytostatic.  
 OS Homo sapiens.

XX

```
PN US2003036635-A1.
XX
PD 20-FEB-2003.
XX
PF 28-AUG-2002; 2002US-00230163.
XX
PR 25-JUL-2000; 2000US-0220638P.
XX
PR 01-JUN-2001; 2001WO-US017800.
XX
PR 29-JUN-2001; 2001WO-US021066.
XX
PR 09-APR-2002; 2002US-00119480.
XX
PA (GETH ) GENENTECH INC.
XX
PI Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
XX
DR WPI; 2003-342045/32.
XX
DR N-PSDB; ACA66888.
XX
PT One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT useful for the manufacture of a medicament for diagnosing or treating
PT tumor.
XX
PS Claim 11; Fig 96; 314pp; English.
XX
CC The present invention relates to the isolation of novel human PRO
CC polypeptides, and the polynucleotide sequences encoding them. The PRO
CC polypeptides are secreted and transmembrane proteins. The PRO
CC polypeptides and polynucleotides are useful for preparing a medicament
CC useful in the diagnosis and treatment of tumours. Anti-PRO antibodies are
CC useful in diagnostic assays for PRO, by detecting its expression in
CC specific cells, tissues or serum, and for affinity purification of PRO
CC from recombinant cell culture or natural sources. AB080739-AB080860
CC represent the human PRO polypeptides of the invention. Note: The sequence
CC data for this patent was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov/psipdsIDEntry.html
XX
SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 6; Length 90;
Best Local Similarity 100.0%; Pred. No. 5.8e-38;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYTPATGPADDEAPDAETTAATAATTAAPTT 60
Db 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYTPATGPADDEAPDAETTAATAATTAAPTT 60

Qy 61 ATTAASTTARKDIPVLKRWGDLNPNRVCP 90
Db 61 ATTAASTTARKDIPVLKRWGDLNPNRVCP 90

RESULT 22
AB033752
ID AB033752 standard; protein; 90 AA.
XX
XX AB033752;
XX
XX 17-SEP-2003 (first entry)
XX
XX Novel human secreted and transmembrane protein PRO1160.
XX
XX Human; secreted and transmembrane protein; PRO; cytostatic;
XX antiarthritic; osteopathic; gene therapy; TNF-Agonist-Alpha;
XX chondrocyte stimulator; pericyte stimulator; fibroblast modulator;
XX pharmaceutical; diagnostic; biosensor; bioindicator; lung tumour;
XX colon tumour; breast tumour; prostate tumour; rectal tumour;
XX liver tumour; bone disorder; cartilage disorder; sports injury;
XX arthritis; wound.
XX
XX Homo sapiens.
XX
XX OS
XX
XX US2003045687-A1.
XX
XX
```

```
XX
PD 06-MAR-2003.
XX
PF 12-AUG-2002; 2002US-00218631.
XX
PR 01-JUN-2001; 2001WO-US017800.
XX
PR 29-JUN-2001; 2001WO-US021066.
XX
PR 09-APR-2002; 2002US-00119480.
XX
PA (GETH ) GENENTECH INC.
XX
PI Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
XX
DR WPI; 2003-512315/48.
XX
DR N-PSDB; ACD68640.
XX
PT New genes, and its encoded secreted and transmembrane polypeptides,
PT useful for stimulating Tumor Necrosis Factor alpha, or chondrocyte or
PT pericyte proliferation, especially for treating lung tumors, arthritis or
PT wounds in a mammal.
XX
PS Claim 11; Fig 96; 314pp; English.
XX
CC The invention describes an isolated nucleic acid molecule comprising a
CC sequence with at least 80% identity to: (a) a nucleotide encoding any of
CC 122 PRO (secreted and transmembrane) polypeptides whose sequences are
CC fully defined in the specification; or (b) any of 122 nucleotide
CC sequences having e.g. 4834, 2504 or 1759 bp fully defined in the
CC specification; or the full length coding sequence of any these 122
CC nucleotide sequences. The PRO polypeptides or polynucleotides are useful
CC as pharmaceuticals, diagnostics, biosensors or bioeffectors. These are
CC particularly useful for detecting tumours (e.g. lung tumour, colon
CC tumour, breast tumour, prostate tumour, rectal tumour, or liver tumour)
CC in a mammal, for stimulating the release of TNF-alpha from human blood,
CC for stimulating the proliferation or differentiation of chondrocyte
CC cells, for stimulating proliferation of pericyte cells, or for modulating
CC normal human dermal fibroblast proliferation. The PRO nucleic acid or
CC polypeptide is also useful for treating tumours or various bone and/or
CC cartilage disorders (e.g. sports injuries or arthritis), or wounds. The
CC PRO polypeptides are useful in drug screening, particularly as targets
CC for therapeutic intervention in these diseases, and in the diagnostic
CC determination of the presence of these diseases. The PRO polypeptides are
CC also useful as molecular weight markers, or for chromosome
CC identification. The PRO genes are useful as hybridisation probes, or for
CC screening libraries of human cDNA, genomic DNA or mRNA. The PRO genes may
CC also be used in gene therapy, particularly for replacing a defective
CC gene. This is the amino acid sequence of a novel human secreted and
CC transmembrane PRO polypeptide
XX
SQ Sequence 90 AA;

Query Match 100.0%; Score 458; DB 6; Length 90;
Best Local Similarity 100.0%; Pred. No. 5.8e-38;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYTPATGPADDEAPDAETTAATAATTAAPTT 60
Db 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYTPATGPADDEAPDAETTAATAATTAAPTT 60

Qy 61 ATTAASTTARKDIPVLKRWGDLNPNRVCP 90
Db 61 ATTAASTTARKDIPVLKRWGDLNPNRVCP 90

RESULT 23
ABU13982
ID ABU13982 standard; protein; 90 AA.
XX
XX AC ABU13982;
XX
XX 26-FEB-2003 (first entry)
XX
XX
```

DE Human PRO1160 polypeptide.  
XX Human; PRO polypeptide; secreted protein; transmembrane protein;  
KW genetic disorder; antibacterial; immunosuppressive.  
XX  
OS Homo sapiens.  
XX US2002103125-A1.  
XX  
XX 01-AUG-2002.  
XX  
XX 20-NOV-2001; 2001US-00989731.  
XX  
PR 16-JUN-1997; 97US-0049787P.  
PR 17-OCT-1997; 97US-0062250P.  
PR 05-NOV-1997; 97WO-US020069.  
PR 12-NOV-1997; 97US-0065186P.  
PR 13-NOV-1997; 97US-0065311P.  
PR 24-NOV-1997; 97US-0066770P.  
PR 25-FEB-1998; 98US-0075945P.  
PR 20-MAR-1998; 98US-0078910P.  
PR 28-APR-1998; 98US-0083322P.  
PR 07-MAY-1998; 98US-0084600P.  
PR 28-MAY-1998; 98US-0087106P.  
PR 02-JUN-1998; 98US-0087607P.  
PR 02-JUN-1998; 98US-0087609P.  
PR 02-JUN-1998; 98US-0087759P.  
PR 03-JUN-1998; 98US-0087827P.  
PR 04-JUN-1998; 98US-0088021P.  
PR 04-JUN-1998; 98US-0088025P.  
PR 04-JUN-1998; 98US-0088026P.  
PR 04-JUN-1998; 98US-0088028P.  
PR 04-JUN-1998; 98US-0088029P.  
PR 04-JUN-1998; 98US-0088030P.  
PR 04-JUN-1998; 98US-0088033P.  
PR 04-JUN-1998; 98US-0088326P.  
PR 05-JUN-1998; 98US-0088167P.  
PR 05-JUN-1998; 98US-0088202P.  
PR 05-JUN-1998; 98US-0088212P.  
PR 05-JUN-1998; 98US-0088217P.  
PR 05-JUN-1998; 98US-0088555P.  
PR 10-JUN-1998; 98US-0088734P.  
PR 10-JUN-1998; 98US-0088738P.  
PR 10-JUN-1998; 98US-0088742P.  
PR 10-JUN-1998; 98US-0088810P.  
PR 10-JUN-1998; 98US-0088824P.  
PR 10-JUN-1998; 98US-0088826P.  
PR 11-JUN-1998; 98US-0088858P.  
PR 11-JUN-1998; 98US-0088861P.  
PR 11-JUN-1998; 98US-0088876P.  
PR 12-JUN-1998; 98US-0089105P.  
PR 16-JUN-1998; 98US-0089440P.  
PR 16-JUN-1998; 98US-0089512P.  
PR 16-JUN-1998; 98US-0089514P.  
PR 17-JUN-1998; 98US-0089532P.  
PR 17-JUN-1998; 98US-0089538P.  
PR 17-JUN-1998; 98US-0089598P.  
PR 17-JUN-1998; 98US-0089599P.  
PR 17-JUN-1998; 98US-0089600P.  
PR 17-JUN-1998; 98US-0089653P.  
PR 18-JUN-1998; 98US-0089801P.  
PR 18-JUN-1998; 98US-0089907P.  
PR 18-JUN-1998; 98US-0089908P.  
PR 16-SEP-1998; 98WO-US019330.  
PR 17-SEP-1998; 98WO-US019437.  
PR 07-OCT-1998; 98WO-US021141.  
PR 01-DEC-1998; 98WO-US021508.  
PR 05-JAN-1999; 99WO-US000106.  
PR 08-MAR-1999; 99WO-US005028.  
PR 02-JUN-1999; 99WO-US012252.  
PR 15-SEP-1999; 99WO-US021090.  
PR 15-SEP-1999; 99WO-US021547.  
PR 30-NOV-1999; 99WO-US028313.

PR 01-DEC-1999; 99WO-US028301.  
PR 01-DEC-1999; 99WO-US028634.  
PR 16-DEC-1999; 99WO-US030095.  
PR 20-DEC-1999; 99WO-US030911.  
PR 06-JAN-2000; 2000WO-US000219.  
PR 06-JAN-2000; 2000WO-US000376.  
PR 11-FEB-2000; 2000WO-US003565.  
PR 18-FEB-2000; 2000WO-US004341.  
PR 22-FEB-2000; 2000WO-US004414.  
PR 24-FEB-2000; 2000WO-US004914.  
PR 24-FEB-2000; 2000WO-US005004.  
PR 02-MAR-2000; 2000WO-US005841.  
PR 10-MAR-2000; 2000WO-US006319.  
PR 15-MAR-2000; 2000WO-US006884.  
PR 20-MAR-2000; 2000WO-US007377.  
PR 30-MAR-2000; 2000WO-US008439.  
PR 15-MAY-2000; 2000WO-US013358.  
PR 17-MAY-2000; 2000WO-US013705.  
PR 22-MAY-2000; 2000WO-US014042.  
PR 30-MAY-2000; 2000WO-US014941.  
PR 02-JUN-2000; 2000WO-US015264.  
PR 28-JUL-2000; 2000WO-US020710.  
PR 11-AUG-2000; 2000WO-US022031.  
PR 23-AUG-2000; 2000WO-US023522.  
PR 24-AUG-2000; 2000WO-US023328.  
PR 08-NOV-2000; 2000WO-US030952.  
PR 01-DEC-2000; 2000WO-US032678.  
PR 28-FEB-2001; 2001WO-US006520.  
PR 01-JUN-2001; 2001WO-US017800.  
PR 20-JUN-2001; 2001WO-US019692.  
PR 29-JUN-2001; 2001WO-US021066.  
PR 09-JUL-2001; 2001WO-US021735.  
PR 28-AUG-2001; 2001US-00941992.  
XX  
XX (GETH ) GENENTECH LTD.  
PA Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
XX Ferrata N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;  
PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;  
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;  
PI Zhang Z;  
XX WPI; 2003-102117/09.  
DR N-PSDB; ABX64199.  
XX  
XX Novel secreted and transmembrane polypeptide for modulating biological  
PT activity of cell expressing the polypeptide, identifying agonists or  
PT antagonists of polypeptide, and as molecular weight markers.  
XX  
PS Claim 12; Fig 282; 649pp; English.  
XX  
XX The present invention relates to the isolation of novel human PRO  
CC polypeptides, and the polynucleotide sequences encoding them. The PRO  
CC polypeptides are secreted and transmembrane proteins. The PRO  
CC polypeptides are useful for detecting other PRO polypeptides, for linking  
CC bioactive molecules to cells expressing PRO polypeptides, for modulating  
CC biological activities of cells expressing PRO polypeptides, and for for  
CC identifying agonists or antagonists. The polynucleotide sequences  
CC encoding PRO polypeptides are useful as hybridisation probes, in  
CC chromosome and gene mapping, in the generation of antisense RNA and DNA,  
CC in the preparation of PRO polypeptides, for generating transgenic animals  
CC or knockout animals, to construct hybridisation probes for mapping the  
CC gene which encodes the PRO polypeptide, and for the genetic analysis of  
CC individuals with genetic disorders, in gene therapy, for chromosome  
CC identification, as chromosome markers, and for generating probes for PCR,  
CC Northern analysis, Southern analysis and Western analysis. ABU13860-  
CC ABU14006 represent the human PRO polypeptides of the invention. Note: The  
CC sequence data for this patent was obtained in electronic format directly  
CC from the USPTO web site at seqdata.uspto.gov/psipsIDEntry.html  
XX  
SQ Sequence 90 AA;

Query Match

100.0%; Score 458; DB 6; Length 90;

```
Best Local Similarity 100.0%; Pred. No. 5.8e-38;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYPATGPADDEAPDAETTTAAATTAAPT 60
    |||||
Db 1 MKFLAVLVLLGVSIFLVSAQNPTTAAADTYPATGPADDEAPDAETTTAAATTAAPT 60
Qy 61 ATTAASTARKDIPVLKRWGDLNGRVCP 90
    |||||
Db 61 ATTAASTARKDIPVLKRWGDLNGRVCP 90

RESULT 24
ABU72567
ID ABU72567 standard; protein; 90 AA.
AC ABU72567;
DT 17-JUN-2003 (first entry)
XX Novel human secreted and transmembrane protein PRO1160.
DE Human; secreted and transmembrane protein; cytotstatic; anti-HIV;
KW viricide; hepatotropic; antiinflammatory; neuroprotective; gene therapy;
KW PRO; pharmaceutical; diagnostic; biosensor; bioreactor; malignancy;
KW cancer; ovarian cancer; colorectal cancer; Kaposi's sarcoma; leukaemia;
KW lymphoma; hepatitis B; multiple sclerosis; Crohn's disease;
KW drug screening.
XX Homo sapiens.
OS
XX
XX US2003003531-A1.
XX 02-JAN-2003.
XX 19-NOV-2001; 2001US-00989734.
XX 16-JUN-1997; 97US-0049787P.
PR 17-OCT-1997; 97US-0062250P.
PR 05-NOV-1997; 97WO-US020069.
PR 12-NOV-1997; 97US-0065186P.
PR 13-NOV-1997; 97US-0065311P.
PR 24-NOV-1997; 97US-0066770P.
PR 25-FEB-1998; 98US-0075945P.
PR 20-MAR-1998; 98US-0078910P.
PR 28-APR-1998; 98US-0083322P.
PR 07-MAY-1998; 98US-0084600P.
PR 28-MAY-1998; 98US-0087106P.
PR 02-JUN-1998; 98US-0087607P.
PR 02-JUN-1998; 98US-0087609P.
PR 03-JUN-1998; 98US-0087759P.
PR 04-JUN-1998; 98US-0088021P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088026P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088030P.
PR 04-JUN-1998; 98US-0088033P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088734P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088742P.
PR 10-JUN-1998; 98US-0088810P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088858P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-00889105P.
PR 16-JUN-1998; 98US-0089440P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089532P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089599P.
PR 17-JUN-1998; 98US-0089600P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089801P.
PR 18-JUN-1998; 98US-0089907P.
PR 18-JUN-1998; 98US-0089908P.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 01-DEC-1998; 98WO-US025108.
PR 05-JAN-1999; 99WO-US000106.
PR 08-MAR-1999; 99WO-US005028.
PR 02-JUN-1999; 99WO-US012252.
PR 15-SEP-1999; 99WO-US021090.
PR 15-SEP-1999; 99WO-US021547.
PR 30-NOV-1999; 99WO-US028313.
PR 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 02-MAR-2000; 2000WO-US005841.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 30-MAR-2000; 2000WO-US008439.
PR 15-MAY-2000; 2000WO-US013358.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 08-NOV-2000; 2000WO-US030952.
PR 01-DEC-2000; 2000WO-US032678.
PR 28-FEB-2001; 2001WO-US006520.
PR 01-JUN-2001; 2001WO-US017800.
PR 20-JUN-2001; 2001WO-US019692.
PR 29-JUN-2001; 2001WO-US021066.
PR 09-JUL-2001; 2001WO-US021735.
PR 28-AUG-2001; 2001US-00941992.
XX (GETH ) GENENTECH INC.
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX WPI; 2003-352829/33.
DR N-PSDB; ACA64421.
XX
XX New genes and secreted and transmembrane polypeptides (e.g. PRO183 or
PT PRO184), useful for treating or diagnosing e.g. ovarian cancer, Kaposi's
PT sarcoma, leukemia, lymphoma, hepatitis B, multiple sclerosis or Crohn's
PT disease.
```

XX  
PS Claim 12; Fig 282; 663pp; English.  
XX  
CC The invention describes a new isolated nucleic acid molecule comprising  
CC the full length coding sequence of the DNA deposited with the American  
CC Type Culture Collection (e.g. ATCC Deposit No. 209621, 552-PTA, 819-PTA,  
CC 209439, 203135, etc); or a sequence with at least 80% identity to a DNA  
CC encoding a PRO polypeptide. The PRO polypeptides or polynucleotides are  
CC useful as pharmaceuticals, diagnostics, biosensors or bioreactors. These  
CC are particularly useful for detecting or treating e.g. malignancies or  
CC cancers (e.g. ovarian cancer, colorectal cancer, Kaposi's sarcoma,  
CC leukemia or lymphoma), hepatitis B, multiple sclerosis, or Crohn's  
CC disease in mammals. The PRO polypeptides are useful in drug screening,  
CC particularly as targets for therapeutic intervention in these diseases,  
CC and in the diagnostic determination of the presence of these diseases.  
CC The PRO polypeptides are also useful as molecular weight markers, or for  
CC chromosome identification. The PRO genes are useful as hybridisation  
CC probes, or for screening libraries of human cDNA, genomic DNA or mRNA.  
CC The PRO genes may also be used in gene therapy, particularly for  
CC replacing a defective gene. This is the amino acid sequence of a novel  
CC human secreted and transmembrane PRO polypeptide  
XX  
SQ Sequence 90 AA;  
  
Query Match 100.0%; Score 458; DB 6; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKFLAVLVLLGVSI FLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
DB 1 MKFLAVLVLLGVSI FLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
  
QY 61 ATTAASTTARKDIPVL PKWVGDL PNRVCP 90  
DB 61 ATTAASTTARKDIPVL PKWVGDL PNRVCP 90  
  
Search completed: September 26, 2005, 08:25:22  
Job time : 106.857 secs

XX  
PS Claim 1; SEQ ID NO 266; 128pp; English.  
XX  
CC The present invention describes a method for assessing whether a patient  
CC is afflicted with breast cancer. The method comprises comparing the level  
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
CC ABR47386 to ABR47632) in a patient sample and the normal level of  
CC expression of the marker in a control non-breast cancer sample, where a  
CC significant increase in the level of expression of the marker in the  
CC patient sample and the normal level is an indication that the patient is  
CC afflicted with breast cancer. The breast cancer associated sequences from  
CC the present invention have cytostatic activities and can be used in gene  
CC therapy. The method is useful for diagnosing and treating breast cancer.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 90 AA;  
  
Query Match 100.0%; Score 458; DB 6; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKFLAVLVLLGVSI FLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
DB 1 MKFLAVLVLLGVSI FLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
  
QY 61 ATTAASTTARKDIPVL PKWVGDL PNRVCP 90  
DB 61 ATTAASTTARKDIPVL PKWVGDL PNRVCP 90  
  
Search completed: September 26, 2005, 08:25:22  
Job time : 106.857 secs

XX  
PS Claim 12; Fig 282; 663pp; English.  
XX  
CC The invention describes a new isolated nucleic acid molecule comprising  
CC the full length coding sequence of the DNA deposited with the American  
CC Type Culture Collection (e.g. ATCC Deposit No. 209621, 552-PTA, 819-PTA,  
CC 209439, 203135, etc); or a sequence with at least 80% identity to a DNA  
CC encoding a PRO polypeptide. The PRO polypeptides or polynucleotides are  
CC useful as pharmaceuticals, diagnostics, biosensors or bioreactors. These  
CC are particularly useful for detecting or treating e.g. malignancies or  
CC cancers (e.g. ovarian cancer, colorectal cancer, Kaposi's sarcoma,  
CC leukemia or lymphoma), hepatitis B, multiple sclerosis, or Crohn's  
CC disease in mammals. The PRO polypeptides are useful in drug screening,  
CC particularly as targets for therapeutic intervention in these diseases,  
CC and in the diagnostic determination of the presence of these diseases.  
CC The PRO polypeptides are also useful as molecular weight markers, or for  
CC chromosome identification. The PRO genes are useful as hybridisation  
CC probes, or for screening libraries of human cDNA, genomic DNA or mRNA.  
CC The PRO genes may also be used in gene therapy, particularly for  
CC replacing a defective gene. This is the amino acid sequence of a novel  
CC human secreted and transmembrane PRO polypeptide  
XX  
SQ Sequence 90 AA;  
  
Query Match 100.0%; Score 458; DB 6; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5.8e-38;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKFLAVLVLLGVSI FLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
DB 1 MKFLAVLVLLGVSI FLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAAPT 60  
  
QY 61 ATTAASTTARKDIPVL PKWVGDL PNRVCP 90  
DB 61 ATTAASTTARKDIPVL PKWVGDL PNRVCP 90  
  
RESULT 25  
ABR47515  
ID ABR47515 standard; protein; 90 AA.  
XX  
AC ABR47515;  
XX  
DT 12-JUN-2003 (first entry)  
XX  
DE Breast cancer associated protein sequence SEQ ID NO:266.  
XX  
KW Human; breast cancer; cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO2003004989-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 21-JUN-2002; 2002WO-US019669.  
XX  
PR 21-JUN-2001; 2001US-0299887P.  
PR 27-JUN-2001; 2001US-0301572P.  
PR 18-JUL-2001; 2001US-0306501P.  
PR 25-SEP-2001; 2001US-0325002P.  
PR 05-MAR-2002; 2002US-0362585P.  
PR 14-MAY-2002; 2002US-0380391P.  
XX  
PA (MILL-) MILLENIUM PHARM INC.  
XX  
PI Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kamatkar S;  
PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;  
PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;  
XX  
DR WFI; 2003-210381/20.  
DR N-PSDB; ACC50211.  
XX

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:51:12 ; Search time 19,1209 Seconds  
(without alignments)  
452.882 Million cell updates/sec

Title: US-09-975-502A-8

Perfect score: 458

Sequence: 1 MKFLAVLVLLGVSLFLVSAQ.....KDIPVLPKWVGDLPNGRVCP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: PIR 79:\*\*

2: PIR2:\*\*

3: PIR3:\*\*

4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112.5	24.6	208	2 B30305	submandibular gland
2	108.5	23.7	164	2 I53641	mucin 5AC - human
3	107	23.4	279	2 S53363	mucin 5AC (clone J
4	104	22.7	105	2 D82734	outer membrane pro
5	101.5	22.2	600	2 S07638	spore coat protein
6	100	21.8	477	2 S53362	mucin 5AC (clone J
7	99	21.6	662	2 A45155	mucin 5AC (clone J
8	97.5	21.3	896	2 S36326	clathrin assembly
9	97.5	21.3	901	2 A44825	phosphoprotein, sy
10	97.5	21.3	915	2 S36327	clathrin assembly
11	96.5	21.1	797	1 VGBEX1	glycoprotein X pre
12	96.5	21.1	866	2 T45462	membrane glycoprot
13	96.5	21.1	867	2 T45463	probable envelope
14	96	21.0	750	2 T42614	probable envelope
15	95.5	20.9	322	2 A53715	apomucin precursor
16	94	20.5	94	2 S53365	mucin 5AC (clone C
17	94	20.5	138	1 WNM516	submandibular gland
18	94	20.5	138	2 A30305	submandibular gland
19	94	20.5	620	2 A70525	hypothetical prote
20	93.5	20.4	91	2 C30305	submandibular gland
21	92.5	20.2	400	1 A28172	spasmolytic precu
22	92.5	20.2	442	1 S11712	transcription init
23	92	20.1	340	2 A35630	regulatory protein
24	91.5	20.1	389	2 T33340	hypothetical prote
25	91.5	20.0	489	1 VGBETA	glycoprotein A - t
26	91.5	20.0	635	2 T75477	hypothetical prote
27	91.5	20.0	1224	2 T14007	microtubule-associ
28	91	19.9	62	2 S53366	mucin 5AC (clone M
29	91	19.9	507	2 T44768	antifreeze glycope

30	91	19.9	909	1	QRXLL2	LDL receptor 2 pre
31	90.5	19.8	190	2	C89560	protein C48B4.12a
32	90.5	19.8	1098	2	T08599	probable transcrip
33	90.5	19.8	1779	2	T31085	xylanase - Caldice
34	89.5	19.5	338	2	S28004	probable cell surf
35	89.5	19.5	386	2	B97212	protein containing
36	89	19.4	91	2	A22592	antifreeze protein
37	89	19.4	352	2	G82990	alginate regulator
38	89	19.4	623	2	F75523	osteoblast specifi
39	89	19.4	825	2	T29634	hypothetical prote
40	88.5	19.3	377	2	A48018	mucin 7 precursor,
41	88	19.2	215	2	S55925	probable arabinoga
42	88	19.2	832	2	JC8051	protein tyrosine p
43	88	19.2	1541	2	T02831	AAA protein L4171.
44	87	19.0	98	2	S53367	mucin 5AC (clone M
45	87	19.0	511	2	T35194	transcription init
46	86.5	18.9	802	2	T21315	hypothetical prote
47	86.5	18.9	885	1	VGBESA	glycoprotein B pre
48	85.5	18.7	105	2	AD2605	conserved hypotet
49	85.5	18.7	105	2	C97387	hypothetical prote
50	85.5	18.7	112	2	S33822	salivary glue prot
51	85.5	18.7	329	2	S38082	pathogenesis-relat
52	85.5	18.7	588	2	JC8021	chitinase (SC 3.2.
53	85.5	18.7	752	2	T34355	hypothetical prote
54	85	18.6	339	2	T25562	hypothetical prote
55	85	18.6	345	2	E88103	protein W10G11.5 l
56	85	18.6	1235	2	T13710	protein-tyrosine k
57	85	18.6	2508	2	S61441	surface-associated
58	84.5	18.4	107	2	S12607	salivary glue prot
59	84.5	18.4	417	2	T20327	hypothetical prote
60	84.5	18.4	525	2	A35596	nuclear pore glyco
61	84.5	18.4	846	2	T21700	hypothetical prote
62	84	18.3	288	2	T21790	hypothetical prote
63	84	18.3	352	2	A36128	regulatory protein
64	84	18.3	801	2	T29018	hypothetical prote
65	83	18.1	168	2	S52994	arabinogalactan-li
66	83	18.1	385	2	T18180	proline-rich prote
67	83	18.1	526	2	A56573	nuclear pore compl
68	83	18.1	540	2	S21825	vicilin-like stora
69	83	18.1	564	2	C84456	hypothetical prote
70	83	18.1	767	1	JU0474	glucan 1,4-alpha-g
71	83	18.1	778	1	ALBYG	mucin 2, intestina
72	83	18.1	1513	2	A54895	glycoprotein gp57-
73	82.5	18.0	523	1	A60408	UL73 glycoprotein
74	82	17.9	138	1	QGBEB2	hypothetical prote
75	82	17.9	632	2	T00084	LDL receptor 1 pre
76	82	17.9	909	1	QRXLL1	nuclear envelope p
77	82	17.9	1199	2	A40670	hypothetical prote
78	82	17.9	1229	2	T25697	neurofascin - chic
79	82	17.9	1272	2	S26180	membrane antigen g
80	81.5	17.8	234	1	QGBB43	hypothetical prote
81	81.5	17.8	248	2	E84500	antifreeze glycopr
82	81.5	17.8	822	2	A38420	hypothetical prote
83	81.5	17.8	839	2	F75518	hypothetical prote
84	81.5	17.8	851	2	T22696	hypothetical prote
85	81.5	17.8	1282	2	T25168	hypothetical prote
86	81	17.7	273	2	T44657	protein gp80 (limp
87	81	17.7	592	2	T34446	hypothetical prote
88	81	17.7	770	2	T22808	hypothetical prote
89	81	17.7	790	2	T34293	hypothetical prote
90	81	17.7	927	2	T24031	probable polyketid
91	81	17.7	3643	2	T36410	AcAGP4 - Arabidops
92	80.5	17.6	135	2	T49996	hypothetical prote
93	80.5	17.6	191	2	D98218	conserved hypotet
94	80.5	17.6	191	2	AG3068	conserved hypotet
95	80.5	17.6	232	1	A60095	larval glue protei
96	80.5	17.6	802	1	A36065	protein-tyrosine-p
97	80.5	17.6	1489	2	T31108	cyst germination s
98	80	17.5	88	2	S02720	outer membrane pro
99	80	17.5	98	2	H81072	hypothetical prote
100	80	17.5	379	2	AE3003	conserved hypotet

ALIGNMENTS

```

RESULT 1
B30305
submandibular gland protein (spot 2) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1989 #sequence_revision 03-Aug-1992 #text_change 09-Jul-2004
C;Accession: B30305
R;Dickinson, D.P.; Mirels, L.; Tabak, L.A.; Gross, K.W.
Mol. Biol. Evol. 6, 80-102, 1989
A;Title: Rapid evolution of variants in a rodent multigene family encoding salivary pro
A;Reference number: A30305; MUID:89158788; PMID:2921944
A;Accession: B30305
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-208 <DIC>
A;Cross-references: UNIPROT:Q62265; GB:M33975; NID:g201046; PIDN:AAA40133.1; PID:g201047
C;Superfamily: submandibular gland 16.5K protein

Query Match      24.6%; Score 112.5; DB 2; Length 208;
Best Local Similarity 44.6%; Pred. No. 0.006;
Matches 33; Conservative 9; Mismatches 23; Indels 9; Gaps 3;

QY 1 MKFLAVLLVGVSIFLVSAQNPTTAAPADTYPATGPDADDEAPDAET--TAAATTATTAAP 58
Db 1 MKFLALLVLLGVSTILVSCDPET---NSTETSGTDEAETNSAETSETADSGGNTSSE 56
QY 59 TTATTAASTTARKD 72
Db 57 TQAD---STNENQD 67

RESULT 2
I53641
mucin 5AC - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: I53641
R;Ho, S.B.; Robertson, A.M.; Shekels, L.L.; Lyftogt, C.T.; Niehans, G.A.; Toribara, N.W.
Gastroenterology 109, 735-747, 1995
A;Title: Expression cloning of Gastric mucin complementary DNA and localization of mucin
A;Reference number: I53641; MUID:95385930; PMID:7657101
A;Accession: I53641
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-164 <RES>
A;Cross-references: UNIPROT:Q14851; GB:L46721; NID:g945218; PIDN:AAC41774.1; PID:g945219
C;Genetics:
A;Gene: GDB:MUC5AC
A;Cross-references: GDB:454136; OMIM:158373
A;Map position: 11p15.5-11p15.5

Query Match      23.7%; Score 108.5; DB 2; Length 164;
Best Local Similarity 38.4%; Pred. No. 0.011;
Matches 24; Conservative 10; Mismatches 26; Indels 3; Gaps 1;

QY 12 VSIFLVSAQNPTTAAPADTYPATGPDADDEAPDAETTAATTAATTAAPTATTAASTTARK 71
Db 16 ISVLTTST---TSASTSTSTSGGTPSPVPTTSTTSAPTSTTSAPTSTTSAPTSTSP 72
QY 72 DIP 74
Db 73 SAP 75

RESULT 3
S53363
mucin 5AC (clone JER58) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S53363
R;Guyonnet-Duport, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buisine, M.P.; Gallieue-
```

```

Biochem. J. 305, 211-219, 1995
A;Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich domai
A;Reference number: S53361; MUID:95126907; PMID:7826332
A;Accession: S53363
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-279 <GUY>
A;Cross-references: UNIPROT:Q14888; EMBL:Z34278; NID:g563376; PIDN:CAA84032.1; PID:g56337
A;Experimental source: clone JER58
C;Genetics:
A;Gene: GDB:MUC5AC
A;Cross-references: GDB:454136; OMIM:158373
A;Map position: 11p15.5-11p15.5
C;Keywords: glycoprotein; tandem repeat

Query Match      23.4%; Score 107; DB 2; Length 279;
Best Local Similarity 44.4%; Pred. No. 0.024;
Matches 24; Conservative 10; Mismatches 18; Indels 2; Gaps 1;

QY 18 SAQNPTTAAPADTYPATGP--ADDEAPDAETTAATTAATTAAPTATTAASTTAA 69
Db 34 SAPKSTTSAASTTSITSGPETTPRPVPTTSTTSPTTSPTTSAPTSTTSATTS 87

RESULT 4
D82734
outer membrane protein H-8 precursor XF1024 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: D82734
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20355717; PMID:10910347
A;Note: for a complete list of authors see reference number AS9328 below
A;Accession: D82734
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-105 <SIM>
A;Cross-references: UNIPROT:Q9PEK4; GB:AE003939; GB:AE003849; NID:g9105949; PIDN:AAF83834
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; AJ
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H.
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Junqueira, M.B.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigre
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins, E.
A;Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
A;Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: AS9328
A;Contents: annotation
C;Genetics:
A;Gene: XF1024

Query Match      22.7%; Score 104; DB 2; Length 105;
Best Local Similarity 32.9%; Pred. No. 0.017;
Matches 26; Conservative 11; Mismatches 32; Indels 10; Gaps 1;

QY 4 LAVLLVGVSIFLVSAQNPTTAAPADTYPATGPADDEA-----PDAETTAATTA 53
Db 6 LLIALAMGATLAACGKTPTPTAPTQDSNPASPAANEQAQAADQAANPPADATPAADT 65
QY 54 TTAAPTATTAASTTARKD 72
Db 66 AAAAATAADAATAATTTTPAD 84

RESULT 5
S07638
```



```

A;Map position: 11p15.5-11p15.5
C;Keywords: glycoprotein; tandem repeat

Query Match          21.8%; Score 100; DB 2; Length 477;
Best Local Similarity 33.8%; Pred. No. 0.17;
Matches 24; Conservative 7; Mismatches 18; Gaps 1;

QY 22 PTTAAADYVPATGPPADD-----EAPDAETAAATTAATTAAPTTATT 63
Db 132 PVTAPSPSGRATSPQTSTSMOKSRRTLLVTTSTTPTQSTTSAPTSTTSAPTSTT 191
QY 64 AASTTARKDIP 74
Db 192 SAPTTSTTSTP 202

RESULT 7
A45155
C;Species: Xenopus laevis (African clawed frog)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Jul-2004
C;Accession: A45155
R;Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A;Title: P-domains as shuffled cysteine-rich modules in integrin-associated protein 180
A;Reference number: A45155; MUID:93077556; PMID:1447205
A;Accession: A45155
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-662 <HAU>
A;Cross-references: UNIPROT:Q05049; GB:L02115; NID:g214147; PIDN:AAA74725.1; P
F;162-202/Domain: trefoil homology <TRF1>
F;307-347/Domain: trefoil homology <TRF2>
F;354-394/Domain: trefoil homology <TRF3>
F;526-566/Domain: trefoil homology <TRF4>
F;573-613/Domain: trefoil homology <TRF5>
F;621-661/Domain: trefoil homology <TRF6>

Query Match          21.6%; Score 99; DB 2; Length 662;
Best Local Similarity 43.7%; Pred. No. 0.28;
Matches 31; Conservative 5; Mismatches 21; Indels 14; Gaps 3;

QY 18 SAQNPTTA---APADYVPATG-----PADDEAPDAETAAATTAATTA---APTATT 63
Db 45 TGEDATTAATAAETAAAGAPTITTTTAPATAAGKAPTAAATAPTAAAGAPTATTATG 104
QY 64 AASTTARKDIP 74
Db 105 KAPATAAAPVP 115

RESULT 8
S36326
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 09-Jul-2004
C;Accession: S36326
R;Morris, S.A.; Schroeder, S.; Plessmann, U.; Weber, K.; Ungewickell, E.
EMBO J. 12, 667-675, 1993
A;Title: Clathrin assembly protein AP180: primary structure, domain organization
A;Reference number: S36326; MUID:93178442; PMID:8440257
A;Accession: S36326
A;Molecule type: mRNA
A;Residues: 1-896 <WOR>
A;Cross-references: UNIPROT:Q05140; EMBL:X68877; NID:g55724; PIDN:CAA48748.1; P
C;Keywords: clathrin binding

Query Match          21.3%; Score 97.5; DB 2; Length 896;
Best Local Similarity 40.3%; Pred. No. 0.51;
Matches 31; Conservative 6; Mismatches 29; Indels 11; Gaps 3;

QY 16 LVSAQNPTTAAP-----ADTYVPATGPADDEAPDA-ETTAATTAATTAATTAATTAAS 66

```



A;Note: ORF71  
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog  
F;558-866/Domain: equine herpesvirus 1 glycoprotein homolog <EHG>

Query Match 21.1%; Score 96.5; DB 2; Length 866;  
Best Local Similarity 43.1%; Pred. No. 0.6;  
Matches 28; Conservative 3; Mismatches 21; Indels 13; Gaps 3;

Qy 18 SAQNPTTAAPAD-----TYPATGPADDEAPDAETAAAT---ATTAAPT---TTA 64  
Db 157 ATSPPTTTPTSTTTTATTTVPATSTTTDTTAAATTTAAATTTAAATTTAAATTTAAATTTA 216  
Qy 65 ASTTA 69  
Db 217 ATTTA 221

RESULT 13  
T45463  
membrane glycoprotein [imported] - equine herpesvirus 1  
C;Species: equine herpesvirus 1  
C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T45463  
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.  
J. Equine Sci. 7, 79-87, 1996  
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine herpesvirus 1  
A;Reference number: 222973  
A;Accession: T45463  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-867 <KIR>  
A;Cross-references: UNIPROT:O39782; EMBL:D88734; PIDN:BAA20038.1  
A;Experimental source: isolate 3F clone; strain BK343  
C;Genetics:  
A;Note: ORF71  
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog

Query Match 21.1%; Score 96.5; DB 2; Length 867;  
Best Local Similarity 43.1%; Pred. No. 0.6;  
Matches 28; Conservative 3; Mismatches 21; Indels 13; Gaps 3;

Qy 18 SAQNPTTAAPAD-----TYPATGPADDEAPDAETAAAT---ATTAAPT---TTA 64  
Db 153 ATSPPTTTPTSTTTTATTTVPATSTTTDTTAAATTTAAATTTAAATTTAAATTTAAATTTA 212  
Qy 65 ASTTA 69  
Db 213 ATTTA 217

RESULT 14  
T42614  
probable envelope protein - equine herpesvirus 4 (strain NS80567)  
C;Species: equine herpesvirus 4  
A;Variety: strain NS80567  
C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T42614  
R;Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.  
J. Gen. Virol. 79, 1197-1203, 1998  
A;Title: The DNA sequence of equine herpesvirus-4.  
A;Reference number: 222173; MUID:98264497; PMID:9603335  
A;Accession: T42614  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-750 <TEL>  
A;Cross-references: UNIPROT:O39307; EMBL:AF030027; NID:g2605950; PIDN:AAC59591.1; PID:g2605950  
A;Experimental source: strain NS80567  
C;Genetics:  
A;Note: 71  
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog

Query Match 21.0%; Score 96; DB 2; Length 750;  
Best Local Similarity 48.9%; Pred. No. 0.57;

Matches 23; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

Qy 23 TTAAPADTYPATGPADDEAPDAETAAATTTAAATTTAAATTTAAATTTAAATTTA 69  
Db 242 TTAATTTAAATTTAAATTTESSEASSTLAATTAATTTADTTADTTADTTA 288

RESULT 15  
A53715  
apomucin precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 09-Jul-2004  
C;Accession: A53715  
R;Albone, E.F.; Hagen, F.K.; VanWyckhuysse, B.C.; Tabak, L.A.  
J. Biol. Chem. 269, 16845-16852, 1994  
A;Title: Molecular cloning of a rat submandibular gland apomucin.  
A;Reference number: A53715; MUID:94266905; PMID:8207007  
A;Accession: A53715  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-322 <ALB>  
A;Cross-references: UNIPROT:Q62605; GB:U03407; NID:g476096; PIDN:AAA20966.1; PID:g476097  
C;Superfamily: hydroxyproline-rich glycoprotein  
C;Keywords: glycoprotein; polymorphism; tandem repeat  
F;1-22/Domain: signal sequence #status predicted <SIG>  
F;23-322/Product: apomucin #status predicted <MAT>

Query Match 20.9%; Score 95.5; DB 2; Length 322;  
Best Local Similarity 41.9%; Pred. No. 0.28;  
Matches 26; Conservative 2; Mismatches 29; Indels 5; Gaps 2;

Qy 22 PTTAAPT---YPATGPADDEAPDAETAAATTTAAATTTAAATTTAAATTTAAATTTA 76  
Db 214 PTTKPTTDSITTPATTKPTTDSITTPATTKPTTDSITTPATTKPTTDSITTPATTKIPTT 273

Qy 77 PK 78  
Db 274 PK 275

RESULT 16  
S53365  
mucin 5AC (clone CEL2) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
C;Accession: S53365  
R;Guyonnet-Duperat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buisine, M.P.; Galiege, J.  
Biochem. J. 305, 211-219, 1995  
A;Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich domain  
A;Reference number: S53361; MUID:95126907; PMID:7826332  
A;Accession: S53365  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-94 <GUY>  
A;Cross-references: UNIPROT:Q14886; EMBL:Z34276; NID:g563372; PIDN:CAA84030.1; PID:g563373  
A;Experimental source: clone CEL2  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994  
C;Genetics:  
A;Gene: GDB:MUC5AC  
A;Cross-references: GDB:454136; OMIM:158373  
A;Map position: 11p15.5-11p15.5  
C;Keywords: glycoprotein; tandem repeat

Query Match 20.5%; Score 94; DB 2; Length 94;  
Best Local Similarity 40.4%; Pred. No. 0.11;  
Matches 21; Conservative 10; Mismatches 15; Indels 6; Gaps 1;

Qy 18 SAQNPTTAAPADTYPATGPADDEAPDAETAAATTTAAATTTAAATTTAAATTTAAATTTA 69  
Db 46 SASTTTSASTTSTTSGPGTTPSP-----VPTTSTTSAPTSTTSASTTS 91

RESULT 17

WMMS16  
submandibular gland 16.5K protein - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 09-Jul-2004  
C/Accession: A03298  
R/Windass, J.D.; Mullins, J.J.; Beecroft, L.J.; George, H.; Meacock, P.A.; Williams, B.R.  
Nucleic Acids Res. 12, 1361-1376, 1984  
A/Title: Molecular cloning of cDNAs from androgen-independent mRNA species of DBA/2 mouse  
A/Reference number: A93503; MUID:84144035; PMID:6546617  
A/Accession: A03298  
A/Molecule type: mRNA  
A/Residues: 1-138 <WIN>  
A/Cross-references: UNIPROT:P02815; GB:X00349; NID:G51367; PIDN:CAA25098.1; PID:G51368  
C/Comment: This protein contains a hydrophobic amino-terminal sequence that is similar to  
C/Superfamily: submandibular gland 16.5K protein  
C/Keywords: Glycoprotein; submandibular gland  
F;25,72,89,94/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 20.5%; Score 94; DB 1; Length 138;  
Best Local Similarity 49.1%; Pred. No. 0.17;  
Matches 26; Conservative 7; Mismatches 14; Indels 6; Gaps 2;

QY 1 MKFLAVLVLLGVSTFLVSAQNPTTAAPADTYPATGPADD--EAPDAETTTAAAT 51  
Db 1 MKFLALLVLVGLVSTILVSCDDPET---NSTETSGTADSAGENTGTQADST 49

RESULT 18  
A30305  
submandibular gland protein (spot 1) precursor - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 22-Nov-1989 #sequence\_revision 03-Aug-1992 #text\_change 09-Jul-2004  
C/Accession: A30305  
R/Dickinson, D.P.; Wierls, L.; Tabak, L.A.; Gross, K.W.  
Mol. Biol. Evol. 6, 80-102, 1989  
A/Title: Rapid evolution of variants in a rodent multigene family encoding salivary pro  
A/Reference number: A30305; MUID:89158788; PMID:2921944  
A/Accession: A30305  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-138 <DIC>  
A/Cross-references: UNIPROT:P02815; GB:M33974; NID:g201044; PIDN:AAA40132.1; PID:g201045  
A/Note: the authors translated the codon AAC for residue 104 as Glu  
C/Superfamily: submandibular gland 16.5K protein

Query Match 20.5%; Score 94; DB 2; Length 138;  
Best Local Similarity 49.1%; Pred. No. 0.17;  
Matches 26; Conservative 7; Mismatches 14; Indels 6; Gaps 2;

QY 1 MKFLAVLVLLGVSTFLVSAQNPTTAAPADTYPATGPADD--EAPDAETTTAAAT 51  
Db 1 MKFLALLVLVGLVSTILVSCDDPET---NSTETSGTADSAGENTGTQADST 49

RESULT 19  
A70525  
hypothetical protein Rv0312 - Mycobacterium tuberculosis (strain H37RV)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C/Accession: A70525  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A/Authors: Squares, R.; Sulatton, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A/Reference number: A70500; MUID:98259587; PMID:9634230  
A/Accession: A70525  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-620 <COL>  
A/Cross-references: UNIPROT:O07239; GB:Z96800; GB:AL123456; NID:g3261800; PIDN:CAB09584.  
A/Experimental source: strain H37RV

```

C:Genetics:
A:Gene: RV0312

Query Match      20.5%; Score 94; DB 2; Length 620;
Best Local Similarity 40.6%; Pred. No. 0.71;
Matches 26; Conservative 3; Mismatches 31; Indels 4; Gaps 2;

Qy 18 SAQNPTTAAAPADTYPATGPADDEAPDAETT---AAATTATTAAAP-TTATTAAASTTARKDI 73
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 512 AAPATTIVAPAPPPPTQVVTTTAPPVTTTPSPPTTTTTPPSTTTTTPPPVTITSTI 571
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 74 PVLP 77
      |||
Db 572 PTIP 575

RESULT 20
C30305
submandibular gland protein (spot 1) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Nov-1989 #sequence_revision 22-Nov-1989 #text_change 09-Jul-2004
C:Accession: C30305
R:Dickinson, D.P.; Mirels, L.; Tabak, L.A.; Gross, K.W.
Mol. Biol. Evol. 6, 80-102, 1989
A:Title: Rapid evolution of variants in a rodent multigene family encoding salivary protein
A:Reference number: A30305; MUID:89158788; PMID:2921944
A:Accession: C30305
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-91 <DIC>
A:Cross-references: UNIPROT:Q63557
C:Superfamily: submandibular gland 16.5K protein

Query Match      20.4%; Score 93.5; DB 2; Length 91;
Best Local Similarity 37.5%; Pred. No. 0.12;
Matches 27; Conservative 7; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MKFLIALVLLGVSLFVSAQNPTTAAAPADTYPATGPADDEAPDAETTAAATTATTAAPTT 60
Db 1 MKFLIALVLLGVSLFVSAQNPTTAAAPADTYPATGPADDEAPDAETTAAATTATTAAPTT 60
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 61 ATTAASTTARKD 72
      |||
Db 60 DAVDGDAPAEQD 71

RESULT 21
A28172
spasmolysin precursor - African clawed frog
N:Alternate names: prospasmolysin
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A28172
R:Hoffmann, W.
J. Biol. Chem. 263, 7686-7690, 1988
A:Title: A new repetitive protein from Xenopus laevis skin highly homologous to pancrat
A:Reference number: A28172; MUID:88227968; PMID:3372504
A:Accession: A28172
A:Molecule type: mRNA
A:Residues: 1-400 <HOF>
A:Cross-references: UNIPROT:P10667; GB:M19971; NID:G214791; PIDN:AAA49960.1; PID:G214792
C:Superfamily: spasmolysin; trefoil homology
C:Keywords: duplication; glycoprotein; skin; tandem repeat
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-400/Product: spasmolysin #status predicted <MAT>
F:21-69/Product: spasmolysin I #status predicted <SP1>
F:23-63/Domain: trefoil homology <TRF1>
F:72-348/Product: spasmolysin-glycopeptide #status predicted <SPG>
F:74-114/Domain: trefoil homology <TRF2>
F:115-309/Region: PEST sequence
F:123-131,231-239/Region: 9-residue repeats (D/E-T-T-T-A-S-T-T-A)
F:132-230,240-248/Region: 9-residue repeats (E-T-T-T-T-V-P-T-T-P)
F:149-252,276-299/Region: 4-residue repeats (E-T-T-T)
F:249-252,276-299/Region: 4-residue repeats (E-T-T-T)

```

RESULT 23  
A35630  
regulatory protein algr3 - Pseudomonas aeruginosa  
C;Species: Pseudomonas aeruginosa  
C;Date: 28-Sep-1990 #sequence\_revision 28-Sep-1990 #text\_change 09-Jul-2004  
C;Accession: A35630  
R;Kato, J.; Mista, T.K.; Chakrabarty, A.M.  
Proc Natl Acad Sci U S A. 87, 2887-2891, 1990  
A>Title: Algr3, a protein resembling eukaryotic histone H1, regulates alginate production in Pseudomonas aeruginosa  
A;Reference number: AJ5630; PMID:90222135; PMID:2109318  
A;Accession: A35630  
A>Status: preliminary; nucleic acid sequence not shown; not compared with cDNA clones from other strains  
A:Molecule type: DNA  
A;Residues: 1-340 <KAT>  
A;Cross-references: UNIPROT:P15276; GB:M35259  
C;Keywords: DNA binding; transcription regulation

Query Match      20.1%; Score 92; DB 2; Length 340;

Best Local Similarity    39.2%; Pred. No. 0.59;

Matches          20; Conservative         7; Mismatches    24; Indels            0; Gaps

Qy    19 AQNPPTAAPADTYPATGPADEAPDAETAAATTATAATTAAPTTATTAASSTA 69  
     ||||| :||| |:: :||||| ::||| :| :  
Db    288 AAQPVAAKPAAKPATAPAANAATPSGATAASAAGSATPAAGSNGCAAPTSS 338  
     ||||| :||| |:: :||||| ::||| :| :

RESULT 24  
T33340  
hypothetical protein K07D4.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T33340  
R;Henkhaus, J.; Wohldmann, P.  
submitted to the EMBL Data Library, July 1998  
A>Description: The sequence of C. elegans cosmid K07D4.  
A;Reference number: Z21327  
A;Accession: T33340  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A;Residues: 1-389 <HEN>  
A;Cross-references: UNIPROT:O76573; EMBL:AF077534; PIDN:AAC26290.1; GSPDB:I-GSPII  
A;Experimental source: strain Bristol N2; clone K07D4  
C;Genetics:  
A;Gene: CESP:K07D4.6  
A;Map position: 2  
A;Introns: 66/2; 119/3; 168/3; 210/2; 236/1

Query Match      20.1%; Score 92; DB 2; Length 389;

Best Local Similarity    36.2%; Pred. No. 0.68;

Matches          21; Conservative         5; Mismatches    28; Indels            4; Gaps

Qy    23 TTAAPADYTPATGPADDEAPDAETAAATTATAATTAAPTTTASTTKKDIFVLPKWV 80  
     ||||| :||| |:: :||||| ::||| :| :  
Db    274 TTTEPTTTTTTTQTOTTTVPSTISTISTSTTTTTTTTTTTSDDL----LMV 332  
     ||||| :||| |:: :||||| ::||| :| :

RESULT 25  
VGbeta  
glycoprotein A - turkey herpesvirus (strain H2)  
C;Species: turkey herpesvirus  
A>Note: host Meleagris gallopavo (common turkey)  
C;Date: 30-Jun-1990 #sequence\_revision 30-Jun-1990 #text\_change 09-Jul-2004  
C;Accession: HQJ162; JS0351  
R;Kato, A.; Sato, I.; Ihara, T.; Ueda, S.; Ishihama, A.; Hirai, K.  
Gene 84, 399-405, 1989  
A>Title: Homology between herpesvirus of turkey and Marek's disease virus t  
A;Reference number: HQJ162; PMID:90128284; PMID:2558972  
A;Accession: HQJ162  
A:Molecule type: DNA  
A;Residues: 1-489 <KAT>  
A;Cross-references: UNIPROT:P18535  
C;Genetics:



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:51:12 ; Search time 19,1209 Seconds  
(without alignments)  
452.882 Million cell updates/sec

Title: US-09-975-502A-6  
Perfect score: 450  
Sequence: 1 MKLSVCLLVTLALCCYQAN.....LQKSLIAEVLVKLKCSV 90

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	206	45.8	111	1 BORT1	prostatic steroid-
2	177	39.3	95	2 S68231	FIG22 protein prec
3	166.5	37.0	98	1 BORT2	prostatic steroid-
4	110	24.4	91	2 JS0036	Clara cell 10K pro
5	102	22.7	91	1 UGRB	uteroglobin precu
6	101	22.4	96	1 UGRB	uteroglobin precu
7	100	22.2	96	2 A36581	polychlorinated bi
8	97	21.6	91	1 UGRBL	uteroglobin precu
9	83.5	18.6	113	2 JC2026	cell specific 10K
10	77.5	17.2	94	2 S17449	probable ligand-bi
11	70	15.6	107	2 JC1127	major allergen cha
12	69	15.3	1623	2 T01369	ABC transporter At
13	66.5	14.8	578	2 T24735	hypothetical prote
14	66.5	14.8	3079	1 RGBY12	probable GTPase-ac
15	66	14.7	102	2 G97516	hypothetical prote
16	64	14.2	25	2 S26651	uteroglobin precu
17	64	14.2	797	2 D86459	probable disease r
18	63	14.0	92	2 JC1136	major allergen cha
19	62.5	13.9	371	2 B88986	protein C50H11.13
20	62.5	13.9	929	2 C90531	cation-transportin
21	62	13.8	109	2 C56413	major allergen Fel
22	62	13.8	284	2 T36313	probable oxidoredu
23	61.5	13.7	200	2 H86556	riboflavin synthas
24	61.5	13.7	200	2 E72066	riboflavin synthas
25	61.5	13.7	416	2 AF2076	hypothetical prote
26	61.5	13.7	1094	2 T00814	RNA-directed DNA p
27	61	13.6	340	2 T46942	SmcC-like regulato
28	61	13.6	812	2 H86265	protein F3F19.18 [
29	60.5	13.4	92	2 A56413	major allergen Fel

30	60.5	13.4	234	2	T31886	hypothetical prote
31	60.5	13.4	315	2	T39444	hypothetical prote
32	60.5	13.4	392	2	A88125	protein f12C9.4 [1
33	60.5	13.4	1390	1	TVHOME	hepatocyte growth
34	60	13.3	100	2	T11037	hypothetical prote
35	60	13.3	123	35	I51604	cholecystokinin pr
36	60	13.3	246	2	T28166	hypothetical prote
37	60	13.3	346	2	T38750	hypothetical prote
38	60	13.3	664	2	C84869	probable receptor
39	59.5	13.2	244	2	AB1120	conserved hypothet
40	59.5	13.2	247	2	AE1480	conserved hypothet
41	59	13.1	261	2	T43579	type III secretion
42	58.5	13.0	961	2	AE2270	exonuclease ABC c
43	58.5	13.0	1097	2	S68685	adenylate cyclase
44	58	12.9	140	2	H64411	hypothetical prote
45	58	12.9	199	2	C75213	hypothetical prote
46	58	12.9	247	1	A64590	probable 3-oxoacyl
47	58	12.9	335	2	T20428	hypothetical prote
48	58	12.9	594	2	E88956	protein ZK697.5 [1
49	58	12.9	724	2	T25700	mechanosensory pro
50	57.5	12.8	95	1	BORT3	prostatic steroid-
51	57.5	12.8	192	2	S52249	response regulator
52	57.5	12.8	314	2	T32293	hypothetical prote
53	57.5	12.8	591	2	H86501	ATP synthase subun
54	57.5	12.8	591	2	B72121	ATP synthase, chai
55	57.5	12.8	805	2	T24399	hypothetical prote
56	57	12.7	163	2	C97308	probable acetyltra
57	57	12.7	282	1	YPD0D1	prestalk D11 prote
58	57	12.7	388	2	T06035	hypothetical prote
59	57	12.7	441	2	T38239	hypothetical prote
60	57	12.7	608	1	ABONS1	serum albumin 1 pr
61	57	12.7	608	1	ABONS2	serum albumin 2 pr
62	57	12.7	1107	2	T15884	hypothetical prote
63	56.5	12.6	137	1	QOBE32	BKRF2 protein - hu
64	56.5	12.6	262	2	H64311	hypothetical prote
65	56.5	12.6	556	2	S02154	NADH2 dehydrogenas
66	56.5	12.6	607	2	T23085	hypothetical prote
67	56.5	12.6	677	2	T23083	hypothetical prote
68	56.5	12.6	838	2	T04449	hypothetical prote
69	56.5	12.6	872	2	T10582	hypothetical prote
70	56.5	12.6	1385	2	S61236	major capsid prote
71	56	12.4	251	2	T32200	hypothetical prote
72	56	12.4	308	2	T12990	hypothetical prote
73	56	12.4	405	2	A55967	1-phosphatidylinos
74	56	12.4	416	2	D96692	hypothetical prote
75	56	12.4	434	2	D71480	probable flagellum
76	56	12.4	446	2	E71635	aspartokinase (lys
77	56	12.4	540	2	C70751	probable fadD10 pr
78	56	12.4	773	2	H96818	hypothetical prote
79	56	12.4	1653	2	G65028	hypothetical prote
80	55.5	12.3	76	2	A63991	DNA-directed RNA p
81	55.5	12.3	88	2	JC1126	major allergen cha
82	55.5	12.3	195	2	JH0680	ciliary neurotroph
83	55.5	12.3	207	2	A83540	phospholipase acce
84	55.5	12.3	309	2	T46226	hypothetical prote
85	55.5	12.3	363	2	F69878	conserved hypothet
86	55.5	12.3	428	2	S56679	mitosis-specific c
87	55.5	12.3	441	2	AF3437	cyclopropane-fatty
88	55.5	12.3	585	2	E85040	hypothetical prote
89	55.5	12.3	616	2	C95861	probable ABC trans
90	55.5	12.3	723	2	E97177	uncharacterized co
91	55.5	12.3	779	2	G87573	xanthine dehydroge
92	55	12.2	209	2	C81691	conserved hypothet
93	55	12.2	314	2	AG3054	succinoglycan bios
94	55	12.2	314	2	F98231	succinoglycan bios
95	55	12.2	345	2	T48313	hypothetical prote
96	55	12.2	429	2	T16656	hypothetical prote
97	55	12.2	465	2	C81130	aspartate ammonia-
98	55	12.2	465	2	G81836	aspartate ammonia-
99	55	12.2	482	2	E70460	flagellar hook pro
100	55	12.2	492	2	A97200	membrane associate

## ALIGNMENTS

```
RESULT 1
BORT1
Prostatic steroid-binding protein chain C1 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change 09-Jul-2004
C:Accession: A92348; A90348; S42615; A03252
R:Parker, M.; Needham, M.; White, R.
Nature 298, 92-94, 1982
A>Title: Prostatic steroid binding protein: gene duplication and steroid binding.
A:Reference number: A92386; MUID:82220075; PMID:6896362
A:Molecule type: mRNA
A:Accession: A92386
A:Residues: 1-111 <PAR>
A:Cross-references: UNIPROT:P02782
R:Liao, S.; Chen, C.; Huang, I.-Y.
J. Biol. Chem. 257, 122-125, 1982
A>Title: Prostate alpha-protein. Complete amino acid sequence of the component that inhibits
A:Reference number: A92348; MUID:82075873; PMID:7198120
A:Molecule type: protein
A:Residues: 24-73, 'D', '75-89, 'E', '91, 'G', '93-111 <LIA>
R:Delaey, B.; Rombauts, W.; Volckaert, G.; Peeters, B.; Mous, J.; Heyns, W.
Biochem. Soc. Trans. 10, 51, 1982
A>Title: Identification of a complementary-DNA clone containing part of the sequence inf
A:Reference number: S42615; MUID:83234456; PMID:6698048
A:Accession: S42615
A:Molecule type: mRNA
A:Residues: 1-3, 'IK', '6-89, 'E', '91, 'G', '93-111 <DE2>
A:Cross-references: EMBL:V01545; NID:g57108; PID:CAA24787.1; PID:g57109
C:Comment: Steroid-binding protein, the principal secretory protein in rat prostatic flu
. The chains of each dimer are linked by disulfide bonds.
C:Superfamily: uteroglobin
C:Keywords: heterotetramer; prostate; steroid binding
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-111/Product: prostatic steroid-binding protein chain C1 #status experimental <MAT>

Query Match 45.8%; Score 206; DB 1; Length 111;
Best Local Similarity 49.5%; Pred. No. 2.1e-16;
Matches 45; Conservative 14; Mismatches 30; Indels 2; Gaps 2;

QY 1 MKLSVCLLLVTLALCCYQAN-AEFCPALVSELLDFFIFSEPLFKLSLAKFDAPPEAAVK 59
Db 4 VELSCLLLIM-LAVCCYEANASQICELVAHETISFLMKSEELKLEWYNAPPAVEAK 62

QY 60 LGVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
Db 63 LEVKRCVDQMSGDLVVAETLVVFLKCDV 93

RESULT 2
S68231
FHG22 protein precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 21-Aug-1998
C:Accession: S68231
R:Dominguez, P.
FEBS Lett. 376, 257-261, 1995
A>Title: Cloning of a Syrian hamster cDNA related to sexual dimorphism: establishment of
A:Reference number: S68231; MUID:96105393; PMID:7498554
A:Accession: S68231
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-95 <DOM>
```

A:Cross-references: EMBL:Z66540

C:Superfamily: uteroglobin

F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-95/Product: FHG22 protein #status predicted <MAT>

Query Match 39.3%; Score 177; DB 2; Length 95;  
Best Local Similarity 43.3%; Pred. No. 3.9e-13;  
Matches 39; Conservative 19; Mismatches 30; Indels 2; Gaps 2;  
QY 1 MKLSVCLLLVTLALCCYQAN-AEFCPALVSELLDFFIFSEPLFKLSLAKFDAPPEAAVK 59  
Db 1 MKLSVCLLLVTLALCCYQAN-AEFCPALVSELLDFFIFSEPLFKLSLAKFDAPPEAAVK 60

QY 60 LGVKRCTDQ-MSLQKRSLIAEVLVKILKCC 88

Db 61 VEVKKCIDSTLNVLEKMEMGKILAEVVGVC 90

## RESULT 3

BORT2

Prostatic steroid-binding protein chain C2 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 15-Oct-1982 #sequence\_revision 15-Oct-1982 #text\_change 09-Jul-2004

C:Accession: A03251; A26671

R:Parker, M.; Needham, M.; White, R.

Nature 298, 92-94, 1982

A>Title: Prostatic steroid binding protein: gene duplication and steroid binding.

A:Reference number: A92386; MUID:82220075; PMID:6896362

A:Accession: A03251

A:Molecule type: mRNA

A:Residues: 1-98 <PAR>

A:Cross-references: UNIPROT:P02781; GB:J00776; NID:G206448; PIDN:AAA51641.1; PID:G206450

R:Delaey, B.; Dirckx, L.; Decourt, J.L.; Claessens, F.; Peeters, B.; Rombauts, W.

Nucleic Acids Res. 15, 1627-1641, 1987

A>Title: Rat prostatic binding protein: the complete sequence of the C2 gene and its flar

A:Reference number: A26671; MUID:87146484; PMID:2881277

A:Accession: A26671

A:Molecule type: DNA

A:Residues: 1-25, 'O', '26-86, 'I', '88-94, 'VLMQINPRGRWFSEIN' <DEL>

A:Cross-references: GB:X05034; NID:956857; PIDN:CAA28708.1; PID:956858

C:Comment: Steroid-binding protein, the principal secretory protein in rat prostatic flu

. The chains of each dimer are linked by disulfide bonds.

C:Superfamily: uteroglobin

C:Keywords: heterotetramer; prostate; steroid binding

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-98/Product: prostatic steroid-binding protein chain C2 #status predicted <MAT>

Query Match 37.0%; Score 166.5; DB 1; Length 98;  
Best Local Similarity 42.6%; Pred. No. 6.4e-12;  
Matches 40; Conservative 12; Mismatches 37; Indels 5; Gaps 2;

QY 1 MKLSVCLLLVTLALCCYQAN----AEFCPALVSELLDFFIFSEPLFKLSLAKFDAPPEAV 56

Db 1 MRLSLCLLTI-LVWCYEAAGQTLAGVCOALQDVITFLNPEELKRELEFADPEAV 59

QY 57 AAKLGVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

Db 60 EANKLVKRCINKIMYGDRLSMGTSLVFTMLKCDV 93

## RESULT 4

JS0036

Clara cell 10K protein precursor - human

N:Alternate names: urinary protein 1

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004

C:Accession: JS0036; PS0309; A56890; I38397

R:Singh, G.; Katyal, S.L.; Brown, W.E.; Phillips, S.; Kennedy, A.L.; Anthony, J.; Squeglia

Biochim. Biophys. Acta 950, 329-337, 1988

A>Title: Amino-acid and cDNA nucleotide sequences of human Clara cell 10KDa protein.

A:Reference number: JS0036; MUID:89000784; PMID:3167058

A:Accession: JS0036

A:Molecule type: mRNA







## RESULT 8

UGRBL  
uteroglobin precursor - brown hare  
N;Alternate names: blastokinin  
C;Species: Lepus capensis (brown hare)  
C;Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jul-2004  
C;Accession: A23825  
R;Lopez de Haro, M.S.; Nieto, A.  
Biochem. J. 235, 895-898, 1986  
A;Title: Nucleotide and derived amino acid sequences of a cDNA coding for pre-uteroglobin  
A;Reference number: A23825; MUID:86323069; PMID:3019311  
A;Accession: A23825  
A;Molecule type: mRNA  
A;Residues: 1-91 <LOP>  
A;Cross-references: UNIPROT:P06913; GB:M25609; NID:g164246; PIDN:AAA30960.1; PID:g164247  
A;Experimental source: lung  
C;Comment: Uteroglobin, synthesized in the uterus and lung, is secreted by the uterus up  
C;Complex: homodimer linked by two disulfide bonds  
C;Superfamily: uteroglobin  
C;Keywords: lung; steroid binding; uterus  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;24-91/Product: uteroglobin #status predicted <MAT>  
F;24/Disulfide bonds: interchain (to 90) #status predicted  
F;90/Disulfide bonds: interchain (to 24) #status predicted

Query Match 21.6%; Score 97; DB 1; Length 91;

Best Local Similarity 30.2%; Pred. No. 0.00058;  
Matches 26; Conservative 15; Mismatches 45; Indels 0; Gaps 0;

Qy 1 MKLSVCLLVTLALCCYQANAEFCPALVSELDFFPFISSEPLFKLSLAKFDAPPRAVAKL 60

Db 1 MKLTITLALVTALLCSPAGICGPFARHVIENLLGTSSYGTSLKEFQPDAMKDAQM 60

Qy 61 GVKRECTDMSLQKRSLIAEVLVKIL 86

Db 61 QMKKVLDTLPQTRENIKLTETIKV 86

## RESULT 9

JC2026  
cell specific 10K protein - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 17-Mar-1999  
C;Accession: JC2026  
R;Ray, M.K.; Magdalen, S.; O'Malley, B.W.; DeMayo, F.J.  
Biochem. Biophys. Res. Commun. 197, 163-171, 1993  
A;Title: Cloning and characterization of the mouse clara cell specific 10 kDa protein gene  
A;Reference number: JC2026; MUID:94071937; PMID:7916613  
A;Accession: JC2026  
A;Molecule type: DNA  
A;Residues: 1-113 <RAY>  
C;Comment: This protein is the major secretory product of the Clara cell and binds to po  
C;Superfamily: uteroglobin  
F;75/Region: ochre stop codon

Query Match 18.6%; Score 83.5; DB 2; Length 113;

Best Local Similarity 23.5%; Pred. No. 0.025;  
Matches 24; Conservative 19; Mismatches 42; Indels 17; Gaps 1;

Qy 1 MKLSVCLLVTLALCCYQ-----ANAEFCPALVSELDFFPFISSEPLFK 43

Db 1 MKIAITVVMLSICCSLLEKLLSCFFFPATPASSDIPCGLQVLEALLMSESGIV 60

Qy 44 LSLAKFDAPPRAVAAGLVKRECTDMSLQKRSLIAEVLVKIL 85

Db 61 ASLAFPPGSLDXNAGTQLKLVTLTQETINIMKLTETIL 102

## RESULT 10

S17449  
probable ligand-binding protein RYD5 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Nov-1999

C;Accession: S17449

R;Dear, T.N.; Boehm, T.; Keverne, E.B.; Rabbitts, T.H.

EMBO J. 10, 2813-2819, 1991

A;Title: Novel genes for potential ligand-binding proteins in subregions of the olfactory

A;Reference number: S17447; MUID:92007724; PMID:1915264

A;Accession: S17449

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-94 <DEA>

A;Cross-references: EMBL:X60661; NID:g57735; PIDN:CAA43068.1; PID:g57736

Query Match 17.2%; Score 77.5; DB 2; Length 94;

Best Local Similarity 28.4%; Pred. No. 0.1;

Matches 25; Conservative 15; Mismatches 45; Indels 3; Gaps 2;

Qy 1 MKLSVCLLVTLALCCYQANAEFCPALVSELDFFPFISSEPLFKLSLAKFDAPPRAVA 57

Db 1 MKGSSALLVALTVLCICGLTRADDNEFFMEFLQTLVGTPEELYEGPLGKYNVNDMAKA 60

Qy 58 AKLGVRKCTDMSLQKRSLIAEVLVKIL 85

Db 61 ALTELKSCIDELQPVHKQLVLLVQVL 88

## RESULT 11

JC1127

major allergen chain 2 precursor, short form - cat

C;Species: Felis silvestris catus (domestic cat)

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004

C;Accession: JC1127

R;Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.  
Gene 113, 263-268, 1992

A;Title: Expression and genomic structure of the genes encoding FdI, the major allergen

A;Reference number: JC1126; MUID:92241678; PMID:1572548

A;Accession: JC1127

A;Molecule type: DNA

A;Residues: 1-107 <GRI>

A;Cross-references: UNIPROT:P30440; GB:X62478; NID:g395406; PIDN:CAA44345.1; PID:g395407

A;Experimental source: skin

C;Genetics:

A;Gene: Ch2

A;Introns: 21/1; 81/3

C;Keywords: glycoprotein

F;1-17/Domain: signal sequence #status predicted <SIG>

F;18-107/Product: major allergen chain 2, short form #status predicted <MAT>

F;50/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.6%; Score 70; DB 2; Length 107;

Best Local Similarity 36.3%; Pred. No. 0.85;

Matches 33; Conservative 8; Mismatches 38; Indels 12; Gaps 5;

Qy 3 LSVCLLVTLALCCYQANAEFCPALVSELDFFPFI-----SEPLFKLSLAKFDPA-PPRAVA 57

Db 5 LLVLALVTQALGVKM--AETCPI----FYDVFFAVANGNELLDLDLSLTKVNATEPERTA 58

Qy 58 AKLGVRKCTDMSLQKRSLIAEVLVKILK 88

Db 59 MK-KIOPCYVENGILSRVLDGLVMIAINEYC 88

## RESULT 12

T01369

ABC transporter AtMRP2 [imported] - Arabidopsis thaliana

N;Alternate names: multidrug-resistance protein homolog T29F13.13

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004

C;Accession: T01369; D84759

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, May 1998

A;Description: Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence.

A;Reference number: Z14179



RESULT 15  
G97516  
hypothetical protein AGR\_C\_2384 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: G97516  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A/Reference number: A97359; MUID:21608551; PMID:11743194  
A/Accession: G97516  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-102 <KUR>  
A/Cross-references: UNIPROT:Q8U5C3; GB:AE007869; PIDN:AAK87088.1; PID:GL15156348; GSPDB:G  
C/Genetics:  
A/Gene: AGR\_C\_2384  
A/Map position: circular chromosome

Query Match 14.7%; Score 66; DB 2; Length 102;  
Best Local Similarity 41.2%; Pred. No. 2.3;  
Matches 21; Conservative 2; Mismatches 24; Indels 4; Gaps 1;  
Qy 12 LALCCYQAN----AEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAA 58  
Db 37 LALSGLANMLRQAVNDPGFLAGLTLFMSHEDLMAPCAATDTPPETVAA 87

RESULT 16  
S26651  
uteroglobin precursor - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 05-Mar-1994 #sequence\_revision 21-Jan-1997 #text\_change 09-Jul-2004  
C/Accession: S26651  
R/Wolf, M.; Klug, J.; Hackenberg, R.; Gessler, M.; Grzeschik, K.H.; Beato, M.; Suske, G.  
Hum. Mol. Genet. 1, 371-378, 1992  
A/Title: Human CC10, the homologue of rabbit uteroglobin: genomic cloning, chromosomal 1  
A/Reference number: S26651; MUID:93250776; PMID:1284526  
A/Accession: S26651  
A/Status: preliminary; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-25 <WOL>  
A/Cross-references: UNIPROT:P11684; EMBL:X59875; NID:G29728; PIDN:CAA42532.1; PID:G29729  
C/Genetics:  
A/Gene: GDB:UGB  
A/Cross-references: GDB:135040; OMIM:192020  
A/Map position: 11q12-11q13  
C/Superfamily: uteroglobin  
C/Keywords: steroid binding; uterus  
F/1-21/Domain: signal sequence #status predicted <SIG>  
F/22-25/Product: uteroglobin (fragment) #status predicted <MAT>

Query Match 14.2%; Score 64; DB 2; Length 25;  
Best Local Similarity 81.2%; Pred. No. 1;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 MKLSVCLLLVTLALCC 16  
Db 1 MKLAVTLVTLALCC 16

RESULT 17  
D86459  
probable disease resistance protein [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: D86459  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: D86459  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-797 <STO>  
A/Cross-references: UNIPROT:Q9FW44; GB:AE005172; NID:G10998939; PIDN:AAAG26078.1; GSPDB:G  
C/Genetics:  
A/Map position: 1

Query Match 14.2%; Score 64; DB 2; Length 797;  
Best Local Similarity 25.2%; Pred. No. 29;  
Matches 27; Conservative 19; Mismatches 35; Indels 26; Gaps 6;  
Qy 1 MKLSVCLLLVTLAL-----CCYQANAEFCPALVSELDFFFISEPLFKL-SLAKFD--- 50  
Db 690 LRLYACPELISLPVEVCELPCLKYVDISQCVSLVS-----LPEKFGKLSLEKIDMRE 742  
Qy 51 ----APPEAVAALGLVKKR--C---TDQMSLQKRSLIAEVLVKILKCC 88  
Db 743 CSLILGLPSSVAALVSLRHVICDEETSSMWMWVKVPFELCIEVAKKC 789

RESULT 18  
JC1136  
major allergen chain 1 precursor A - cat  
C/Species: Felis silvestris catus (domestic cat)  
C/Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C/Accession: JC1136  
R/Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.  
Gene 113, 263-268, 1992  
A/Title: Expression and genomic structure of the genes encoding FdI, the major allergen  
A/Reference number: JC1126; MUID:92241678; PMID:1572548  
A/Accession: JC1136  
A/Molecule type: DNA  
A/Residues: 1-92 <GRI>  
A/Cross-references: UNIPROT:P30438  
C/Genetics:  
A/Gene: Chi  
A/Introns: 21/1; 83/3  
F/1-22/Domain: signal  
F/23-92/Product: major allergen chain 1 #status predicted <MAT>

Query Match 14.0%; Score 63; DB 2; Length 92;  
Best Local Similarity 29.9%; Pred. No. 4.7;  
Matches 26; Conservative 15; Mismatches 42; Indels 4; Gaps 4;  
Qy 1 MKLSVCLLLVTLALCCYQ-ANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAA 58  
Db 1 MKGACVLVLLWALLISGGNCEICPA-VKRDVDLFLTGTPDEYVEQVQYNALPVVLEN 59  
Qy 59 KLGVKRCITD-QMSLQKRSLIAEVLVKI 84  
Db 60 ARILKNCVDKAMTEEDKENALSVLDKI 86

RESULT 19  
E88986  
protein C50H11.13 [imported] - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C/Accession: E88986  
R/anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A/Reference number: A75000; MUID:99069613; PMID:9851916  
A/Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_ele  
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A/Accession: E88986

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-371 <STO>  
A;Cross-references: UNIPROT:O16474; GB:chr\_V; PIDN:AAB65974.1; PID:g2315473; GSPDB:GN000  
C;Genetics:  
A;Gene: C50H11.13  
A;Map position: 5

Query Match 13.9%; Score 62.5; DB 2; Length 371;  
Best Local Similarity 30.8%; Pred. No. 20;  
Matches 20; Conservative 8; Mismatches 32; Indels 5; Gaps 1;

Qy 1 MKSVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAK-----FDAPPEA 55  
Db 43 LTLPIILLTCYILCRSKNAHFLPYLCSLAANFVLLSTIFLSVLAKNTDLVVDITPGF 102  
Qy 56 VAAKL 60  
Db 103 LVCKI 107

RESULT 20  
C90531  
cation-transporting p-type atpase [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C;Species: Mycoplasma pulmonis  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C;Accession: C90531  
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A;Reference number: A99512; MUID:21267165; PMID:11353084  
A;Accession: C90531  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-929 <KUR>  
A;Cross-references: UNIPROT:Q98R55; GB:AL445566; PID:g14089568; PIDN:CAC13328.1; GSPDB:G  
A;Experimental source: strain UAB CTIP  
C;Genetics:  
A;Gene: MYPV 1550  
A;Genetic code: SGC3  
C;Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain

Query Match 13.9%; Score 62.5; DB 2; Length 929;  
Best Local Similarity 29.2%; Pred. No. 49;  
Matches 26; Conservative 13; Mismatches 27; Indels 23; Gaps 4;

Qy 3 LSVCLLLVT-----LALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVAA 58  
Db 258 LGISLFLITFIQTALANFASQK-----IYINSLIVAIISLS-VAAIPEGLAA 304  
Qy 59 -----KLGVKRCCTDOMSLQKRSLIAEVL 81  
Db 305 FTTIITSLGVKRWKSKNALVKSLLAVEAL 333

RESULT 21  
C56413  
major allergen Fel d1 chain 2 precursor - cat  
C;Species: Felis silvestris catus (domestic cat)  
C;Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 09-Jul-2004  
C;Accession: C56413; JCI1145  
R;Morgenstern, J.P.; Griffith, I.J.; Brauer, A.W.; Rogers, B.L.; Bond, J.F.; Chapman, M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 9690-9694, 1991  
A;Title: Amino acid sequence of Fel d1, the major allergen of the domestic cat: protein  
A;Reference number: A56413; MUID:92052157; PMID:1946388  
A;Accession: C56413  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-109 <MOR>  
A;Cross-references: UNIPROT:P30440; GB:W77341; NID:g163822; PIDN:AAC41616.1; PID:g163823  
R;Griffith, I.J.; Craig, S.; Folllock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.  
Gene 113, 263-268, 1992  
A;Title: Expression and genomic structure of the genes encoding Fd1, the major allergen

A;Reference number: JCI126; MUID:92241678; PMID:1572548  
A;Accession: JCI145  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 18-109 <GRI>  
A;Experimental source: salivary gland  
C;Keywords: glycoprotein  
F;1-17/Domain: signal sequence #status predicted <SIG>  
F;50/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.8%; Score 62; DB 2; Length 109;  
Best Local Similarity 36.8%; Pred. No. 7.2;  
Matches 32; Conservative 6; Mismatches 37; Indels 12; Gaps 5;

Qy 3 LSVCLLLVTALCCYQANAEFCPALVSELDFFFI-----SEPLFKLSLAKFDA-PPEAVA 57  
Db 5 LLVLALLVTQALGVKM--AETCPI----FYDFFFAVANGNEILLDLSLTKVNATEPERTA 58  
Qy 58 AKLGVKRCCTDOMSLQKRSLIAEVLVKI 84  
Db 59 MK-KIQDCYVENGLISRVLGLVMTTI 84

RESULT 22  
T36313  
probable oxidoreductase - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Mar-2003  
R;Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, March 1999  
A;Reference number: Z21604  
A;Accession: T36313  
A;Status: preliminary; translated from GB/EMBL/DBBJ  
A;Molecule type: DNA  
A;Residues: 1-284 <SAU>  
A;Cross-references: EMBL:AL035654; PIDN:CAB38592.1; GSPDB:GN00070; SCOEDB:SCB8.14C  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCB8.14C  
C;Superfamily: prephenate dehydrogenase/arogenate dehydrogenase, feedback inhibition-ins

Query Match 13.8%; Score 62; DB 2; Length 284;  
Best Local Similarity 31.3%; Pred. No. 18;  
Matches 26; Conservative 8; Mismatches 43; Indels 6; Gaps 1;

Qy 4 SVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAVA----- 57  
Db 15 SVGLALRRHGVTYLTLDADPEAALAGERCAGFAAPPQADIAVLAVPPDQVAPVLAEH 74  
Qy 58 AKLGVKRCCTDOMSLQKRSLIAEV 80  
Db 75 QKLGATRCYTDVSGVKVRLHREV 97

RESULT 23  
H86556  
riboflavin synthase [imported] - Chlamydomophila pneumoniae (strain J138)  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: H86556  
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise  
Nucleic Acids Res. 28, 2311-2314, 2000  
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A;Reference number: A86491; MUID:20330349; PMID:10871362  
A;Accession: H86556  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-200 <STO>  
A;Cross-references: UNIPROT:Q9Z820; GB:BA000008; NID:g9878903; PIDN:BAA98738.1; GSPDB:GN  
A;Experimental source: strain J138  
C;Genetics:  
A;Gene: ribC

C;Superfamily: riboflavin synthase, alpha subunit

Query Match 13.7%; Score 61.5; DB 2; Length 200;  
Best Local Similarity 52.0%; Pred. No. 15;  
Matches 13; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

Qy 49 FDAPPEAVA-AKLGVKRCTQMSIQ 72  
Db 61 FDIPIETLACTTLGKRCSDQVNL 85

# RESULT 24

E72066

riboflavin synthase, alpha chain CP0220 [imported] - Chlamydothila pneumoniae (strains C

C;Species: Chlamydothila pneumoniae, Chlamydia pneumoniae

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004

C;Accession: E72066; D81600

R;Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A;Reference number: A72000; MUID:99206606; PMID:10192388

A;Accession: E72066

A;Molecule type: DNA

A;Residues: 1-200 <ARN>

A;Cross-references: UNIPROT:Q92820; GB:AE001638; GB:AE001363; NID:94376819; PIDN:AA01867

A;Experimental source: strain CWL029

R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A;Reference number: A81500; MUID:20150255; PMID:10684935

A;Accession: D81600

A;Molecule type: DNA

A;Residues: 1-200 <REA>

A;Cross-references: GB:AE002183; GB:AE002161; NID:G7189146; PIDN:AAF38089.1; PID:G718915

A;Experimental source: strain AR39, HL cells

C;Genetics:

A;Gene: ribC; CP0220

C;Superfamily: riboflavin synthase, alpha subunit

Query Match 13.7%; Score 61.5; DB 2; Length 200;  
Best Local Similarity 52.0%; Pred. No. 15;  
Matches 13; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

Qy 49 FDAPPEAVA-AKLGVKRCTQMSIQ 72  
Db 61 FDIPIETLACTTLGKRCSDQVNL 85

# RESULT 25

AF2070

hypothetical protein all2116 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C;Accession: AF2070

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AF2070

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-416 <KUR>

A;Cross-references: UNIPROT:Q8YV65; GB:BA000019; PIDN:BA073815.1; PID:G17131207; GSPDB:C

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all2116

Query Match 13.7%; Score 61.5; DB 2; Length 416;  
Best Local Similarity 25.9%; Pred. No. 30;  
Matches 28; Conservative 14; Mismatches 39; Indels 27; Gaps 4;

Qy 3 LSVCLLLVTLALCCYQANAEFCP--ALVSELLDFFFI---SEPLF-----KLSLAKFD-- 50  
Db 27 LMMALSVVATVWLSFPLNAQITPSAALASELRSGLVTLNIDSDVLFERDRUKTSLQKLDKL 86  
Qy 51 -----APPEAVA-AKLGVKRCTQMSLQKRSLIAEVLVK 83  
Db 87 NFNTVYPVAVNMGYITLYPSKVAKVIGRAIDPTFGLQGRDMLKEIVTE 134

Search completed: September 26, 2005, 08:31:10

Job time : 22.1209 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:56:52 ; Search time 25.0549 Seconds  
(without alignments)  
268.147 Million cell updates/sec

Title: US-09-975-502A-8

Perfect score: 458

Sequence: 1 MKELAVLVLLGVSLFLVSAQ.....KDIPVLPKWVGDLPNGRVCP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*

2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*

3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*

4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*

5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	458	100.0	90	3	US-09-247-155-86
2	458	100.0	90	4	US-09-602-877A-98
3	107	23.4	786	3	US-09-103-429A-3
4	107	23.4	788	4	US-09-294-663-3
5	107	23.4	805	3	US-09-103-429A-4
6	107	23.4	807	4	US-09-294-663-4
7	101	22.1	853	4	US-09-489-039A-11009
8	96.5	21.1	162	4	US-09-270-767-42570
9	94	20.5	166	4	US-09-270-767-37485
10	92.5	20.2	215	3	US-09-188-930-347
11	92.5	20.2	215	4	US-09-312-283C-347
12	92.5	20.2	215	4	US-09-188-930-195
13	92.5	20.2	237	3	US-09-188-930-338
14	92.5	20.2	237	4	US-09-312-283C-195
15	92.5	20.2	237	4	US-09-312-283C-338
16	92.5	20.2	258	4	US-09-252-991A-31620
17	92.5	20.2	442	1	US-08-363-255-4
18	92.5	20.2	442	1	US-08-363-255-11
19	92.5	20.2	442	1	US-08-363-255-11
20	91.5	20.0	178	4	US-09-252-991A-20280
21	91.5	20.0	489	2	US-08-663-566A-6
22	91.5	20.0	489	2	US-08-023-610-6
23	91.5	20.0	489	2	US-08-288-065A-6
24	91.5	20.0	489	2	US-08-362-240A-6
25	91.5	20.0	489	3	US-08-804-372A-4
26	91.5	20.0	489	5	PCT-US95-10245-6
27	91	19.9	357	4	US-09-949-016-6409

28	91	19.9	359	4	US-09-949-016-9131	Sequence 9131, Ap
29	89.5	19.5	338	4	US-09-538-092-144	Sequence 144, Ap
30	89.5	19.5	941	4	US-07-757-022B-14	Sequence 14, Appl
31	89.5	19.5	1022	4	US-07-757-022B-84	Sequence 84, Appl
32	89.5	19.5	1038	4	US-07-757-022B-74	Sequence 74, Appl
33	89.5	19.5	1049	4	US-07-757-022B-58	Sequence 58, Appl
34	89.5	19.5	1140	4	US-07-757-022B-104	Sequence 104, Appl
35	89.5	19.5	1270	4	US-07-757-022B-44	Sequence 44, Appl
36	89.5	19.5	1311	4	US-07-757-022B-42	Sequence 42, Appl
37	89.5	19.5	1313	4	US-07-757-022B-142	Sequence 142, Appl
38	89.5	19.5	1314	4	US-07-757-022B-50	Sequence 50, Appl
39	89.5	19.5	1320	4	US-07-757-022B-46	Sequence 46, Appl
40	89.5	19.5	1320	4	US-07-757-022B-60	Sequence 60, Appl
41	89.5	19.5	1320	4	US-10-164-595-58	Sequence 58, Appl
42	89.5	19.5	1354	4	US-07-757-022B-48	Sequence 48, Appl
43	89.5	19.5	1361	4	US-07-757-022B-40	Sequence 40, Appl
44	89.5	19.5	1363	4	US-07-757-022B-52	Sequence 52, Appl
45	89.5	19.5	1404	4	US-07-757-022B-2	Sequence 2, Appl
46	89.5	19.5	1404	4	US-07-757-022B-62	Sequence 62, Appl
47	89.5	19.5	1404	4	US-10-164-595-78	Sequence 78, Appl
48	89.5	19.5	1404	4	US-09-298-970A-1	Sequence 1, Appl
49	89.5	19.5	1411	4	US-09-949-016-10827	Sequence 10827, A
50	89	19.4	399	4	US-09-252-991A-22853	Sequence 22853, A
51	88.5	19.3	128	4	US-09-270-767-38128	Sequence 38128, A
52	88.5	19.3	128	4	US-09-270-767-53345	Sequence 53345, A
53	86.5	18.9	885	1	US-08-042-747A-8	Sequence 8, Appl
54	86.5	18.9	885	3	US-08-804-439A-23	Sequence 23, Appl
55	86.5	18.9	885	3	US-08-720-229-23	Sequence 23, Appl
56	86	18.8	208	4	US-09-252-991A-25785	Sequence 25785, A
57	86	18.8	626	4	US-09-248-796A-27023	Sequence 27023, A
58	85.5	18.7	155	4	US-09-252-991A-25452	Sequence 25452, A
59	85.5	18.7	1686	4	US-09-902-540-15792	Sequence 15792, A
60	84.5	18.4	154	4	US-09-270-767-33789	Sequence 33789, A
61	84.5	18.4	154	4	US-09-270-767-49006	Sequence 49006, A
62	84.5	18.4	161	4	US-09-252-991A-30841	Sequence 30841, A
63	84	18.3	216	4	US-09-248-796A-25076	Sequence 25076, A
64	83	18.1	806	1	US-08-270-076A-11	Sequence 11, Appl
65	82.5	18.0	186	4	US-09-540-236-2748	Sequence 2748, Ap
66	82	17.9	57	4	US-09-060-767B-6	Sequence 6, Appl
67	81.5	17.8	403	4	US-09-248-796A-15512	Sequence 15512, A
68	81	17.7	305	4	US-09-107-532A-5258	Sequence 5258, Ap
69	80.5	17.6	139	4	US-09-148-545-160	Sequence 160, App
70	80.5	17.6	202	4	US-09-252-991A-32054	Sequence 32054, A
71	80.5	17.6	385	4	US-09-248-796A-14515	Sequence 14515, A
72	80.5	17.6	477	4	US-09-248-796A-21985	Sequence 21985, A
73	80.5	17.6	802	1	US-08-015-985-1	Sequence 1, Appl
74	80.5	17.6	802	4	US-09-280-597-1	Sequence 1, Appl
75	80.5	17.6	807	4	US-09-949-016-7356	Sequence 7356, Ap
76	80	17.5	149	4	US-09-270-767-4367	Sequence 4367, A
77	80	17.5	332	4	US-09-489-039A-11558	Sequence 11558, A
78	80	17.5	362	1	US-08-415-751-5	Sequence 5, Appl
79	80	17.5	674	4	US-09-949-016-7196	Sequence 7196, Ap
80	80	17.5	2972	3	US-09-579-181-2	Sequence 2, Appl
81	80	17.5	3118	3	US-09-579-181-1	Sequence 1, Appl
82	79.5	17.4	189	4	US-09-270-767-39457	Sequence 39457, A
83	79.5	17.4	189	4	US-08-928-361B-8	Sequence 8, Appl
84	79.5	17.4	216	3	US-08-928-361B-8	Sequence 8, Appl
85	79.5	17.4	216	4	US-09-588-995A-8	Sequence 8, Appl
86	79.5	17.4	357	1	US-08-078-683A-8	Sequence 8, Appl
87	79.5	17.4	357	4	US-08-471-970A-8	Sequence 8, Appl
88	79.5	17.4	357	4	US-09-723-677B-8	Sequence 8, Appl
89	79.5	17.4	1837	3	US-08-928-361B-5	Sequence 5, Appl
90	79.5	17.4	1837	4	US-09-588-995A-5	Sequence 5, Appl
91	79	17.2	88	4	US-09-489-039A-12377	Sequence 12377, A
92	79	17.2	175	3	US-08-700-651-12	Sequence 12, Appl
93	79	17.2	175	3	US-08-928-361B-17	Sequence 17, Appl
94	79	17.2	175	4	US-09-588-995A-17	Sequence 17, Appl
95	79	17.2	354	4	US-09-902-540-15411	Sequence 15411, A
96	79	17.2	494	4	US-09-248-796A-16546	Sequence 16546, A
97	79	17.2	516	4	US-09-248-796A-19517	Sequence 19517, A
98	78.5	17.1	174	4	US-09-248-796A-25211	Sequence 25211, A
99	78.5	17.1	178	4	US-09-248-796A-21678	Sequence 21678, A
100	78.5	17.1	249	3	US-08-700-651-15	Sequence 15, Appl







```
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52702
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-52702

Query Match      20.5%; Score 94; DB 4; Length 166;
Best Local Similarity 38.9%; Pred. No. 0.017;
Matches 28; Conservative 13; Mismatches 25; Indels 6; Gaps 3;

QY 1 MKFLAVLVLG-VSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETAAATTAATTAAPT 59
Db 87 MKYSCVLLLLATVACFLVSLSSATT--TTTIDATTTTTTTASSSDT--TTTSSSDT 141
QY 60 TATTAASTTARK 71
Db 142 TTTTEASSSKKK 153

RESULT 11
US-09-188-930-347
; Sequence 347, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 347
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Mouse
US-09-188-930-347

Query Match      20.2%; Score 92.5; DB 3; Length 215;
Best Local Similarity 41.4%; Pred. No. 0.033;
Matches 24; Conservative 3; Mismatches 26; Indels 5; Gaps 1;

QY 18 SAQNPN-----TTAAPADTYPATGPADDEAPDAETAAATTAATTAAPTAASTTAR 70
Db 152 SASTPIPESSETQTINTPTVTAKTAKDTANTTAVTTANTTANTTAVTTAKTTAK 209

RESULT 12
US-09-312-283C-347
; Sequence 347, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0

Query Match      20.2%; Score 92.5; DB 3; Length 215;
Best Local Similarity 41.4%; Pred. No. 0.037;
Matches 24; Conservative 3; Mismatches 26; Indels 5; Gaps 1;

QY 18 SAQNPN-----TTAAPADTYPATGPADDEAPDAETAAATTAATTAAPTAASTTAR 70
Db 152 SASTPIPESSETQTINTPTVTAKTAKDTANTTAVTTANTTANTTAVTTAKTTAK 209

RESULT 13
US-09-188-930-195
; Sequence 195, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 195
; LENGTH: 237
; TYPE: PRT
; ORGANISM: mouse
US-09-188-930-195

Query Match      20.2%; Score 92.5; DB 3; Length 237;
Best Local Similarity 41.4%; Pred. No. 0.037;
Matches 24; Conservative 3; Mismatches 26; Indels 5; Gaps 1;

QY 18 SAQNPN-----TTAAPADTYPATGPADDEAPDAETAAATTAATTAAPTAASTTAR 70
Db 152 SASTPIPESSETQTINTPTVTAKTAKDTANTTAVTTANTTANTTAVTTAKTTAK 209

RESULT 14
US-09-188-930-338
; Sequence 338, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Mouse
US-09-188-930-338

Query Match      20.2%; Score 92.5; DB 3; Length 237;
Best Local Similarity 41.4%; Pred. No. 0.037;
Matches 24; Conservative 3; Mismatches 26; Indels 5; Gaps 1;
```



; SEQUENCE CHARACTERISTICS:  
; LENGTH: 442 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-363-255-4

Query Match 20.2%; Score 92.5; DB 1; Length 442;  
Best Local Similarity 40.4%; Pred. No. 0.079;  
Matches 36; Conservative 7; Mismatches 23; Indels 23; Gaps 6;

QY 19 AQNPTT-----AAPADTYPATGPA--DDEAPDAETTTAAATTA--TTAAPTATTAAAS- 66  
Db 29 AANPVTSRKATAPAPA--APATEPAAVEEAPAKAAAKTTAKKATATAKTTAKKAAAK 86

QY 67 -TTARKDIPVL-----PKWVGDLPLNG 86  
Db 87 KTTAKKGELLEDTEATEEPKAAATEEP 115

## RESULT 19

US-08-363-255-11  
; Sequence 11, Application US/08363255  
; Patent No. 5783386  
; GENERAL INFORMATION:  
; APPLICANT: JACOBS, JR., WILLIAM R.  
; APPLICANT: BLOOM, BARRY R.  
; APPLICANT: COLLINS, DESMOND M.  
; APPLICANT: de LISLE, GEOFFREY W.  
; APPLICANT: PASCOBELLA, LISA  
; APPLICANT: KAWAKAMI, RIKU P.  
; TITLE OF INVENTION: MYCOBACTERIA VIRULENCE FACTORS AND A  
; TITLE OF INVENTION: NOVEL METHOD FOR THEIR IDENTIFICATION  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/363,255  
; FILING DATE: 23-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MONROY, GLADYS H.  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 25237-20002.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 442 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-363-255-11

Query Match 20.2%; Score 92.5; DB 1; Length 442;  
Best Local Similarity 40.4%; Pred. No. 0.079;  
Matches 36; Conservative 7; Mismatches 23; Indels 23; Gaps 6;

QY 19 AQNPTT-----AAPADTYPATGPA--DDEAPDAETTTAAATTA--TTAAPTATTAAAS- 66  
Db 29 AANPVTSRKATAPAPA--APATEPAAVEEAPAKAAAKTTAKKATATAKTTAKKAAAK 86

QY 67 -TTARKDIPVL-----PKWVGDLPLNG 86  
Db 87 KTTAKKGELLEDTEATEEPKAAATEEP 115

## RESULT 20

US-09-252-991A-20280  
; Sequence 20280, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20280  
; LENGTH: 178  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20280

## Query Match

20.0%; Score 91.5; DB 4; Length 178;  
Best Local Similarity 42.5%; Pred. No. 0.033;  
Matches 31; Conservative 5; Mismatches 30; Indels 7; Gaps 3;

QY 24 TAAPADTYPATGPADDEAPDAET---TAAATTATTAA---PTTAT-TAASTTARKDIPVL 76  
Db 106 SATATSKPATSCATGPAPTAAASVPAPAAATTATPAAAPKPTRSTSTIASTSAIGSPSPA 165

QY 77 PKWVGDLPLNGRVC 89

Db 166 SATRRSIPNRRTC 178

## RESULT 21

US-08-663-566A-6  
; Sequence 6, Application US/08663566A  
; Patent No. 5853733  
; GENERAL INFORMATION:

; APPLICANT: Cochran, Mark D  
; APPLICANT: Macdonald, Richard D  
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys  
; TITLE OF INVENTION: and Uses Thereof  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: John P. White  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/663,566A  
; FILING DATE: June 13, 1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P

; REGISTRATION NUMBER: 28,678

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)278-0400

; TELEFAX: (212)391-0526

```
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 489 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-663-566A-6

Query Match      20.0%; Score 91.5; DB 2; Length 489;
Best Local Similarity 34.3%; Pred. No. 0.11;
Matches 34; Conservative 13; Mismatches 35; Indels 17; Gaps 5;

QY 1 MKFLAVLVLLG-VSIFLVSAQNPTTAA-----PADTY-----PATGPAD-----DEAPDAE 45
Db 5 MRVLRLRLTGWVGIFVLVLSLQQTSCAGLPHNVDTTHILTFNPSPISADGVPLSEVPNSP 64

QY 46 TTAATAATTAT--TAAPTATTAASTTARKDIPVLPKWVG 82
Db 65 TTELSTTVATKTAVPTTESTSSSEAHNRSSHKIPDIICD 103

RESULT 22
US-08-023-610-6
; Sequence 6, Application US/08023610
; Patent No. 5928648
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D., Mark D
; APPLICANT: Macdonald Ph.D., Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
; TITLE OF INVENTION: and Uses Thereof
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/023,610
; FILING DATE: February 26, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 489 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-023-610-6

Query Match      20.0%; Score 91.5; DB 2; Length 489;
Best Local Similarity 34.3%; Pred. No. 0.11;
Matches 34; Conservative 13; Mismatches 35; Indels 17; Gaps 5;

QY 1 MKFLAVLVLLG-VSIFLVSAQNPTTAA-----PADTY-----PATGPAD-----DEAPDAE 45
Db 5 MRVLRLRLTGWVGIFVLVLSLQQTSCAGLPHNVDTTHILTFNPSPISADGVPLSEVPNSP 64

QY 46 TTAATAATTAT--TAAPTATTAASTTARKDIPVLPKWVG 82
Db 65 TTELSTTVATKTAVPTTESTSSSEAHNRSSHKIPDIICD 103

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 489 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-023-610-6

Query Match      20.0%; Score 91.5; DB 2; Length 489;
Best Local Similarity 34.3%; Pred. No. 0.11;
Matches 34; Conservative 13; Mismatches 35; Indels 17; Gaps 5;

QY 1 MKFLAVLVLLG-VSIFLVSAQNPTTAA-----PADTY-----PATGPAD-----DEAPDAE 45
Db 5 MRVLRLRLTGWVGIFVLVLSLQQTSCAGLPHNVDTTHILTFNPSPISADGVPLSEVPNSP 64

QY 46 TTAATAATTAT--TAAPTATTAASTTARKDIPVLPKWVG 82
Db 65 TTELSTTVATKTAVPTTESTSSSEAHNRSSHKIPDIICD 103
```

```
RESULT 23
US-08-288-065A-6
; Sequence 6, Application US/08288065A
; Patent No. 5961982
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Macdonald, Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys S-
; TITLE OF INVENTION: HVT-050 and Uses Thereof
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,065A
; FILING DATE: Aug-09-94
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0526
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 489 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-288-065A-6

Query Match      20.0%; Score 91.5; DB 2; Length 489;
Best Local Similarity 34.3%; Pred. No. 0.11;
Matches 34; Conservative 13; Mismatches 35; Indels 17; Gaps 5;

QY 1 MKFLAVLVLLG-VSIFLVSAQNPTTAA-----PADTY-----PATGPAD-----DEAPDAE 45
Db 5 MRVLRLRLTGWVGIFVLVLSLQQTSCAGLPHNVDTTHILTFNPSPISADGVPLSEVPNSP 64

QY 46 TTAATAATTAT--TAAPTATTAASTTARKDIPVLPKWVG 82
Db 65 TTELSTTVATKTAVPTTESTSSSEAHNRSSHKIPDIICD 103

RESULT 24
US-08-362-240A-6
; Sequence 6, Application US/08362240A
; Patent No. 5965138
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Junker, David
; APPLICANT: Wild, Martha A
; TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
```



```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-804-372A-4

Query Match      20.0%; Score 91.5; DB 3; Length 489;
Best Local Similarity 34.3%; Pred. NO. 0.11;
Matches 34; Conservative 13; Mismatches 35; Indels 17; Gaps 5;

QY      1 MKFVLVLVLG-VSIFLVAQNPTTAA-----PATGPAD----DEAPDAE 45
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5 MRVLRVLRGTGWGIFLVLVLQQTSCAGLPHNVDTTHILTFNPSISADGVPLSEVNSP 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      46 TTAATATAT--TAAPTATATAATARKDIPVLKMGVD 82
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      65 TTSLTIVATKTAVPTESTSSSEAHNSSHKIPDIIC 103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: September 26, 2005, 08:32:30
Job time : 27.0549 secs

```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:56:52 ; Search time 25.8901 Seconds  
(without alignments)  
268.147 Million cell updates/sec

Title: US-09-975-502A-5

Perfect score: 475

Sequence: 1 MKLLMVLMLAALSOHCYAGS.....LSNVEVFMQLIYDSSLCDLIF 93

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A COMB.pep:\*

2: /cgn2\_6/ptodata/1/iaa/5B COMB.pep:\*

3: /cgn2\_6/ptodata/1/iaa/6A COMB.pep:\*

4: /cgn2\_6/ptodata/1/iaa/6B COMB.pep:\*

5: /cgn2\_6/ptodata/1/iaa/PGTUS COMB.pep:\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	100.0	93	1	US-08-455-896-2
2	475	100.0	93	2	US-08-933-149-2
3	475	100.0	93	2	US-09-082-343-2
4	475	100.0	93	3	US-09-082-253-2
5	475	100.0	93	3	US-09-215-818-5
6	475	100.0	93	4	US-09-467-602A-5
7	475	100.0	93	4	US-09-162-622-2
8	475	100.0	93	4	US-09-509-015-2
9	475	100.0	93	4	US-08-834-759-503
10	475	100.0	93	5	PCT-US96-08235-2
11	475	100.0	113	4	US-09-949-016-10479
12	475	100.0	410	4	US-08-620-405B-495
13	475	100.0	410	4	US-08-834-759-495
14	475	100.0	743	4	US-08-620-405B-494
15	475	100.0	743	4	US-08-834-759-494
16	475	100.0	1095	4	US-08-620-405B-493
17	475	100.0	1095	4	US-08-834-759-493
18	380	80.0	74	4	US-09-162-622-17
19	281	59.2	95	3	US-08-821-451A-6
20	281	59.2	95	3	US-09-263-810-6
21	281	59.2	95	3	US-09-583-169-6
22	281	59.2	95	4	US-09-471-276-1555
23	281	59.2	108	4	US-08-673-395A-238
24	281	59.2	116	4	US-09-949-016-8424
25	204.5	43.1	95	1	US-08-455-896-7
26	204.5	43.1	95	2	US-08-933-149-7
27	204.5	43.1	95	2	US-09-082-343-7
28	204.5	43.1	29	28	US-09-082-253-7
29	204.5	43.1	29	29	US-09-162-622-7
30	204.5	43.1	31	29	US-09-509-015-7
31	204.5	43.1	31	31	PCT-US96-08235-7
32	196.5	41.4	95	3	US-08-821-451A-27
33	196.5	41.4	95	3	US-09-263-810-27
34	196.5	41.4	95	3	US-09-583-169-27
35	114	24.0	21	4	US-09-834-759-498
36	103	21.7	20	4	US-09-834-759-496
37	102	21.5	20	4	US-09-834-759-499
38	83	17.5	16	4	US-08-933-149-14
39	83	17.5	16	4	US-09-162-622-14
40	83	17.5	16	4	US-09-509-015-14
41	83	17.5	16	5	PCT-US96-08235-14
42	79	16.6	15	4	US-09-834-759-497
43	74	15.6	226	4	US-09-328-352-7363
44	73.5	15.5	513	4	US-09-134-000C-5326
45	73.5	15.5	5024	4	US-09-710-279-2964
46	72	15.2	91	1	US-08-455-896-8
47	72	15.2	91	2	US-08-933-149-8
48	72	15.2	91	2	US-09-082-343-8
49	72	15.2	91	3	US-09-082-253-8
50	72	15.2	91	4	US-08-987-505-1
51	72	15.2	91	4	US-09-162-622-8
52	72	15.2	91	4	US-09-509-015-8
53	72	15.2	91	5	PCT-US96-08235-8
54	72	15.2	94	4	US-09-949-016-10599
55	69.5	14.6	322	4	US-09-248-796A-23680
56	69.5	14.6	347	3	US-09-134-001C-3752
57	69.5	14.6	347	4	US-09-710-279-106
58	69.5	14.6	817	4	US-09-543-681A-4637
59	69	14.5	13	4	US-09-834-759-502
60	69	14.5	155	3	US-08-336-553A-2
61	69	14.5	155	4	US-08-439-157-2
62	69	14.5	155	4	US-09-437-895-2
63	68.5	14.4	302	4	US-09-543-681A-5445
64	68.5	14.4	364	4	US-09-417-485D-42
65	66.5	14.0	136	4	US-09-270-767-32107
66	66.5	14.0	136	4	US-09-270-767-47324
67	66.5	14.0	445	3	US-09-134-001C-3507
68	66.5	14.0	680	4	US-09-583-110-3191
69	66.5	14.0	690	4	US-09-107-433-2835
70	66.5	14.0	10182	3	US-09-134-001C-3159
71	66	13.9	13	4	US-09-834-759-501
72	66	13.9	1786	3	US-08-973-462-8
73	65.5	13.8	333	4	US-09-710-279-1960
74	65.5	13.8	369	3	US-09-134-001C-5149
75	65.5	13.8	649	3	US-09-300-909-13
76	65.5	13.8	649	3	US-09-300-909-26
77	65	13.7	101	4	US-09-252-991A-26467
78	65	13.7	125	4	US-09-248-796A-19285
79	65	13.7	4872	4	US-09-424-783-3
80	64	13.5	194	4	US-09-270-767-47197
81	64	13.5	3033	1	US-07-925-695-9
82	63	13.3	229	4	US-09-489-039A-14255
83	63	13.3	343	4	US-09-270-767-39334
84	63	13.3	343	4	US-09-270-767-54551
85	63	13.3	433	4	US-09-107-532A-6785
86	63	13.3	476	4	US-09-529-157-4
87	63	13.3	630	3	US-08-973-462-9
88	63	13.3	820	4	US-09-248-796A-19901
89	62.5	13.2	218	4	US-09-248-796A-15000
90	62	13.1	229	4	US-09-543-681A-5405
91	62	13.1	541	4	US-09-134-000C-5420
92	62	13.1	894	1	US-08-117-362-4
93	62	13.1	894	1	US-08-486-924-4
94	62	13.1	894	4	US-08-486-929A-4
95	62	13.1	2555	4	US-09-538-092-820
96	61.5	12.9	590	4	US-09-248-796A-18990
97	61.5	12.9	649	3	US-09-300-909-17
98	61.5	12.9	649	3	US-09-300-909-25
99	61.5	12.9	899	4	US-09-107-532A-4503
100	61.5	12.9	945	4	US-09-198-452A-1030

## ALIGNMENTS

```
RESULT 1
US-08-455-896-2
; Sequence 2, Application US/08455896
; Patent No. 5668267
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,896
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-6092
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-455-896-2

Query Match 100.0%; Score 475; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 4.2e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSQHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATNAID 60
Db 1 MKLLMVLMLAALSQHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATNAID 60

QY 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93
Db 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93

RESULT 2
US-08-933-149-2
; Sequence 2, Application US/08933149
; Patent No. 5922836
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
; TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
```

```
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,149
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: HENDERSON, MELODIE W.
; REGISTRATION NUMBER: 37,848
; REFERENCE/DOCKET NUMBER: 6029-6040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-933-149-2

Query Match 100.0%; Score 475; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 4.2e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSQHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATNAID 60
Db 1 MKLLMVLMLAALSQHCYAGSGCPLENNVSKTINPQVSKTEYKELLQEFIDNATNAID 60

QY 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93
Db 61 ELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF 93

RESULT 3
US-09-082-343-2
; Sequence 2, Application US/09082343
; Patent No. 5968754
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,343
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE:
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-082-343-2

Query Match 100.0%; Score 475; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 4.2e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 M K L I V M L M L A A L S Q H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60
Db 1 M K L I V M L M L A A L S Q H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60

Qy 61 E L K E C F L N Q T D E T L S N V E V F M Q L I Y D S S L C D L F 93
Db 61 E L K E C F L N Q T D E T L S N V E V F M Q L I Y D S S L C D L F 93

RESULT 4
US-09-082-253-2
; Sequence 2, Application US/09082253
; Patent No. 6004756
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,253
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

```
; HYPOTHETICAL: NO
; US-09-082-253-2

Query Match 100.0%; Score 475; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 4.2e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 M K L I V M L M L A A L S Q H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60
Db 1 M K L I V M L M L A A L S Q H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60

Qy 61 E L K E C F L N Q T D E T L S N V E V F M Q L I Y D S S L C D L F 93
Db 61 E L K E C F L N Q T D E T L S N V E V F M Q L I Y D S S L C D L F 93

RESULT 5
US-09-215-818-5
; Sequence 5, Application US/09215818A
; Patent No. 6379671
; GENERAL INFORMATION:
; APPLICANT: Colpitts, Tracey
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
; TITLE OF INVENTION: DETECTING DISEASES OF THE BREAST
; FILE REFERENCE: 5972 US.P2
; CURRENT APPLICATION NUMBER: US/09/215,818A
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/912,276
; EARLIER FILING DATE: 1997-08-17
; EARLIER APPLICATION NUMBER: 08/697,105
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: 08/912,149
; EARLIER FILING DATE: 1997-08-15
; EARLIER APPLICATION NUMBER: 08/697,106
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-09-215-818-5

Query Match 100.0%; Score 475; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 4.2e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 M K L I V M L M L A A L S Q H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60
Db 1 M K L I V M L M L A A L S Q H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N A T T N A I D 60

Qy 61 E L K E C F L N Q T D E T L S N V E V F M Q L I Y D S S L C D L F 93
Db 61 E L K E C F L N Q T D E T L S N V E V F M Q L I Y D S S L C D L F 93

RESULT 6
US-09-467-602A-5
; Sequence 5, Application US/09467602A
; Patent No. 6552164
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Colpitts, Tracey L.
; APPLICANT: Russell, John C.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
; TITLE OF INVENTION: DETECTING DISEASES OF THE BREAST
; FILE REFERENCE: 5972 US.P5
; CURRENT APPLICATION NUMBER: US/09/467,602A
; CURRENT FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 08/215,818
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: US 08/912,276
; PRIOR FILING DATE: 1997-08-17
```

```

; PRIOR APPLICATION NUMBER: US 08/697,105
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: US 08/912,149
; PRIOR FILING DATE: 1997-08-15
; PRIOR APPLICATION NUMBER: US 08/697,106
; PRIOR FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-467-602A-5

```

	Query Match	100.0%; Score 475; DB 4; Length 93;
	Best Local Similarity	100.0%; Pred. No. 4.2e-47;
Matches	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	MKLLWVLMLAALSQHCYAGSGCPLENVISKTINPQVKTEYKELLQEFDIDNATNAID 60       
Dd	1	MKLLWVLMLAALSQHCYAGSGCPLENVISKTINPQVKTEYKELLQEFDIDNATNAID 60       
Qy	61	ELKECFINQTDTLSNVVEFMQLIYDSSLCOLF 93     
Dd	61	ELKECFINQTDITLSNVVEFMQLIYDSSLCOLF 93     

```

RESULT 7
US-09-162-622-2
; Sequence 2, Application US/09162622
; Patent No. 6566072
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A
; APPLICANT: FLEMING, TIMOTHY P
; TITLE OF INVENTION: Mammaglobin, A Secreted Mammary-Specific Breast Cancer
; TITLE OF INVENTION: Protein
; FILE REFERENCE: 6029-5134
; CURRENT APPLICATION NUMBER: US/09/162,622
; CURRENT FILING DATE: 1998-09-29
; EARLIER APPLICATION NUMBER: 08/933,149
; EARLIER FILING DATE: 1997-09-18
; EARLIER APPLICATION NUMBER: PCT/US96/08235
; EARLIER FILING DATE: 1996-05-31
; EARLIER APPLICATION NUMBER: 08/455,896
; EARLIER FILING DATE: 1995-05-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-162-622-2

```

	Query Match	100.0%;	Score 475;	DB 4;	Length 93;
	Best Local Similarity	100.0%;	Pred. No.	4.2e-47;	
Matches	93;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
Qy	1	MKLLMWLMLAALSGHCYAGSGCPLENNVISKTNPQVSKEYKELLQEFDIDDNATTNAID	60		
Dd	1	MKLLMWLMLAALSGHCYAGSGCPLENNVISKTNPQVSKEYKELLQEFDIDDNATTNAID	60		
Qy	61	ELKECFINQTDITLSNVVEFMQLIYYDSSLCDLP	93		
Dd	61	ELKECFINQTDITLSNVVEFMQLIYYDSSLCDLP	93		

RESULT 8  
US-09-509-015-2  
; Sequence 2, Application US/09509015  
; Patent No. 6677428  
; GENERAL INFORMATION:  
; APPLICANT: WATSON, MARK S.; FLEMING, TIMOTHY P.  
; TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED

```

; MAMMARY SPECIFIC BREAST CANCER PROTEIN
;
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/509,015
; FILING DATE: 30-May-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/17991
; FILING DATE: 1998-09-18
; APPLICATION NUMBER: 08/933,149
; FILING DATE: 1997-09-18
;
; ATTORNEY/AGENT INFORMATION:
; NAME: KASTEN, DANIEL S.
; REGISTRATION NUMBER: 45,363
; REFERENCE/DOCKET NUMBER: 6029-3654
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
;
; US-09-509-015-2
;
;
; Query Match 100.0%; Score 475; DB 4; Length 93;
; Best Local Similarity 100.0%; Pred. No. 4.2e-47;
; Matches 93; Conservative 0; Mismatches 0; Indels 0;
;
; QY 1 M K L L V I M L M A L S O H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N
;
; Db 1 M K L L V I M L M A L S O H C Y A G S G C P L L E N V I S K T I N P Q V S K T E Y K E L L Q E F I D D N
;
; QY 61 E L K E C F L N O T D E T L S N V E V F M Q L I Y D S S L C D L F 93
;
; Db 61 E L K E C F L N O T D E T L S N V E V F M Q L I Y D S S L C D L F 93
;
; RESULT 9
; US-09-834-759-503
; Sequence 503, Application US/09834759
; Patent No. 6680197
;
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
;
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: Fast-SEQ for Windows Version 3.0

```

```
; SEQ ID NO 503
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-834-759-503

Query Match      100.0%; Score 475; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 4.2e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 60
   |||||||
Db 1 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 60
   |||||||

QY 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93
   |||||||
Db 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93
   |||||||

RESULT 10
PCT-US96-08235-2
; Sequence 2, Application PC/TUS9608235
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08235
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 964796
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
PCT-US96-08235-2

Query Match      100.0%; Score 475; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 4.2e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 60
   |||||||
Db 1 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 60
   |||||||

QY 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93
   |||||||
Db 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93
   |||||||

; SEQ ID NO 10479
; Sequence 495, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10479
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10479

Query Match      100.0%; Score 475; DB 4; Length 113;
Best Local Similarity 100.0%; Pred. No. 5.4e-47;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 60
   |||||||
Db 21 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 80
   |||||||

QY 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93
   |||||||
Db 81 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 113
   |||||||

RESULT 12
US-09-620-405B-495
; Sequence 495, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 495
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-620-405B-495

Query Match      100.0%; Score 475; DB 4; Length 410;
Best Local Similarity 100.0%; Pred. No. 2.9e-46;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 60
   |||||||
Db 1 MKLLVLMMLAALSOHCYAGSGCPLLENVSKTINPQVSKTEYKELLQEFIDNNATTNAID 60
   |||||||

QY 61 ELKECFNLQTDETLSNVVEFMQLIYDSSLCDLF 93
   |||||||
```

```
Db 61 ELKECFLNQTDETLSNVEVFMQLIYDSSSLCDLF 93

RESULT 13
US-09-834-759-495
; Sequence 495, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 495
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-834-759-495

Query Match 100.0%; Score 475; DB 4; Length 410;
Best Local Similarity 100.0%; Pred. No. 2.9e-46;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSOHCYAGSGCPLEENVISKTNPQVSKTEYKELLQEFIDNATNAID 60
Db 1 MKLLMVLMLAALSOHCYAGSGCPLEENVISKTNPQVSKTEYKELLQEFIDNATNAID 60

QY 61 ELKECFLNQTDETLSNVEVFMQLIYDSSSLCDLF 93
Db 61 ELKECFLNQTDETLSNVEVFMQLIYDSSSLCDLF 93

RESULT 14
US-09-620-405B-494
; Sequence 494, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 494
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(743)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-620-405B-494

Query Match 100.0%; Score 475; DB 4; Length 743;
Best Local Similarity 100.0%; Pred. No. 6.3e-46;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSOHCYAGSGCPLEENVISKTNPQVSKTEYKELLQEFIDNATNAID 60
Db 1 MKLLMVLMLAALSOHCYAGSGCPLEENVISKTNPQVSKTEYKELLQEFIDNATNAID 60

QY 61 ELKECFLNQTDETLSNVEVFMQLIYDSSSLCDLF 93
Db 61 ELKECFLNQTDETLSNVEVFMQLIYDSSSLCDLF 93

RESULT 15
US-09-834-759-494
; Sequence 494, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 494
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(743)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-834-759-494

Query Match 100.0%; Score 475; DB 4; Length 743;
Best Local Similarity 100.0%; Pred. No. 6.3e-46;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMVLMLAALSOHCYAGSGCPLEENVISKTNPQVSKTEYKELLQEFIDNATNAID 60
Db 1 MKLLMVLMLAALSOHCYAGSGCPLEENVISKTNPQVSKTEYKELLQEFIDNATNAID 60

QY 61 ELKECFLNQTDETLSNVEVFMQLIYDSSSLCDLF 93
Db 61 ELKECFLNQTDETLSNVEVFMQLIYDSSSLCDLF 93

RESULT 16
US-09-620-405B-493
; Sequence 493, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 493
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
```



```
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(1095)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-620-405B-493

Query Match      100.0%; Score 475; DB 4; Length 1095;
Best Local Similarity 100.0%; Pred. No. 1.1e-45;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLVLMALALSHQCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNATTNAID 60
Db 1 MLLVLMALALSHQCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNATTNAID 60

Qy 61 ELKECFINQDTLSNVVEFMQLIYDSSLCDLF 93
Db 61 ELKECFINQDTLSNVVEFMQLIYDSSLCDLF 93

RESULT 17
US-09-834-759-493
; Sequence 493, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 493
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(1095)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-834-759-493

Query Match      100.0%; Score 475; DB 4; Length 1095;
Best Local Similarity 100.0%; Pred. No. 1.1e-45;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLVLMALALSHQCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNATTNAID 60
Db 1 MLLVLMALALSHQCYAGSGCPLENNVSKTINPQVSKTEYKELQEFIDNATTNAID 60

Qy 61 ELKECFINQDTLSNVVEFMQLIYDSSLCDLF 93
Db 61 ELKECFINQDTLSNVVEFMQLIYDSSLCDLF 93

RESULT 18
US-09-162-622-17
; Sequence 17, Application US/09162622
; Patent No. 6566072
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A
; APPLICANT: FLEMING, TIMOTHY P
; TITLE OF INVENTION: Mammaglobin, A Secreted Mammary-Specific Breast Cancer
; FILE OF INVENTION: Protein
; FILE REFERENCE: 6029-5134
; CURRENT APPLICATION NUMBER: US/09/162,622
; CURRENT FILING DATE: 1998-09-29

; EARLIER APPLICATION NUMBER: 08/933,149
; EARLIER FILING DATE: 1997-09-18
; EARLIER APPLICATION NUMBER: PCT/US96/08235
; EARLIER FILING DATE: 1996-05-31
; EARLIER APPLICATION NUMBER: 08/455,896
; EARLIER FILING DATE: 1995-05-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-162-622-17

Query Match      80.0%; Score 380; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 2.5e-36;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 SGCPLLENNVSKTINPQVSKTEYKELQEFIDNATTNAIDELKECFINQDTLSNV 79
Db 1 SGCPLLENNVSKTINPQVSKTEYKELQEFIDNATTNAIDELKECFINQDTLSNV 60

Qy 80 FMQLIYDSSLCDLF 93
Db 61 FMQLIYDSSLCDLF 74

RESULT 19
US-08-821-451A-6
; Sequence 6, Application US/08821451A
; Patent No. 6068724
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,451A
; FILING DATE: March 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,724
; FILING DATE: March 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-821-451A-6

Query Match      59.2%; Score 281; DB 3; Length 95;
```

```

Best Local Similarity   58.7%; Pred. No. 7.9e-25;
Matches    54; Conservative    13; Mismatches    25; Indels      0; Gaps      0;

Qy  1 MKLLMWLMLAALSQHCVAGSGCPLENNVSKTINPQVSKTEYKELLOEFIDDDNATTNAID 60
     |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  1 MKLLMWLMLAALLHCHVADSGCKLLEDWVKETINSDISIPEYKELLQEFIDSDAAEAMG 60

Qy  61 ELKECFLNQTDETLSNVVEFMQLIYDSSLCDL 92
     :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  61 KFKQCFLNQSHRTLKNFGLMHMTVYDSIWGNM 92
     :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 20
US-09-263-810-6
; Sequence 6, Application US/09263810
; Patent No. 6174992
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; BINDING FACTOR: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,810
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,451
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-263-810-6

Query Match          59.2%; Score 281; DB 3; Length 95;
Best Local Similarity 58.7%; Pred. No. 7.9e-25;
Matches    54; Conservative    13; Mismatches    25; Indels      0; Gaps      0;

Qy  1 MKLLMWLMLAALSQHCVAGSGCPLENNVSKTINPQVSKTEYKELLOEFIDDDNATTNAID 60
     |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  1 MKLLMWLMLAALLHCHVADSGCKLLEDWVKETINSDISIPEYKELLQEFIDSDAAEAMG 60

Qy  61 ELKECFLNQTDETLSNVVEFMQLIYDSSLCDL 92
     :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  61 KFKQCFLNQSHRTLKNFGLMHMTVYDSIWGNM 92
     :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 21
US-09-583-169-6
; Sequence 6, Application US/09583169
```



```

Qy 1 MKLLWLMLAALSQHCHYA-GSCGPLELNVISKTIINPQISKTEYKELLOEFDIDNATNAI 59
Db 1 MKLLVFLFLVTIIPICISYAGSGGSIIDVIRGTINSTVTLHDYMKLVKPYVDHFTTEKAV 60

Qy 60 DEUKCEFLNQDTETLNVEVFMQLIYDSSLG 90
Db 61 KQPKQCFLDQDQTKTLENVGMMEAFNFSSEC 91

```

Search completed: September 26, 2005, 08:32:27  
Job time : 27.8901 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:56:52 ; Search time 25.0549 Seconds  
(without alignments)  
268.147 Million cell updates/sec

Title: US-09-975-502A-6  
Perfect score: 450  
Sequence: 1 MKLSVCLLVTLALCCVQAN.....LQKRSLIAEVLVILKKCSV 90

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	450	100.0	90	3	US-08-821-451A-4
2	450	100.0	90	3	US-09-263-810-4
3	450	100.0	90	3	US-08-912-276-15
4	450	100.0	90	3	US-09-583-169-4
5	450	100.0	90	3	US-09-215-818-6
6	450	100.0	90	4	US-09-467-602A-6
7	450	100.0	90	4	US-09-431-384B-20
8	342	76.0	69	3	US-08-912-276-23
9	342	76.0	69	4	US-09-431-384B-28
10	277	61.6	90	3	US-08-821-451A-2
11	277	61.6	90	3	US-09-263-810-2
12	277	61.6	90	3	US-09-583-169-2
13	267	59.3	53	4	US-09-513-999C-4555
14	267	59.3	53	4	US-09-471-276-841
15	244	54.2	79	4	US-09-673-395A-200
16	219	48.7	45	3	US-08-912-276-22
17	219	48.7	45	4	US-09-431-384B-27
18	202	44.9	90	3	US-08-821-451A-25
19	202	44.9	90	3	US-08-821-451A-26
20	202	44.9	90	3	US-09-263-810-25
21	202	44.9	90	3	US-09-263-810-26
22	202	44.9	90	3	US-09-583-169-25
23	202	44.9	90	3	US-09-583-169-26
24	110	24.4	91	1	US-08-455-896-8
25	110	24.4	91	2	US-08-933-149-8
26	110	24.4	91	2	US-09-082-343-8
27	110	24.4	91	3	US-09-082-253-8
28	110	24.4	91	4	US-08-987-505-1
29	110	24.4	91	4	US-09-162-622-8
30	110	24.4	91	4	US-09-509-015-8
31	110	24.4	91	5	PCT-US96-08235-8
32	110	24.4	94	4	US-09-949-016-10599
33	101	22.4	22	3	US-08-912-276-21
34	101	22.4	22	4	US-09-431-384B-26
35	80	17.8	15	3	US-08-912-276-16
36	80	17.8	15	4	US-09-431-384B-21
37	75.5	16.8	93	2	US-08-964-725-14
38	75.5	16.8	93	4	US-10-081-17A-32
39	75	16.7	15	3	US-08-912-276-19
40	75	16.7	15	4	US-09-431-384B-24
41	73	16.2	15	4	US-09-431-384B-23
42	73	16.2	16	3	US-08-912-276-18
43	72	16.0	15	4	US-09-431-384B-22
44	72	16.0	16	3	US-08-912-276-17
45	70	15.6	109	1	US-07-662-193-5
46	70	15.6	109	3	US-08-300-928C-8
47	70	15.6	109	3	US-08-430-944D-8
48	70	15.6	109	3	US-08-430-014-8
49	70	15.6	109	3	US-08-431-184-8
50	69	15.3	1621	3	US-08-972-927-3
51	66.5	14.8	224	1	US-08-287-959-6
52	66.5	14.8	3079	5	PCT-US94-00198-4
53	66	14.7	15	4	US-09-431-384B-25
54	66	14.7	16	3	US-08-912-276-20
55	63	14.0	15	4	US-09-431-384B-29
56	62	13.8	109	1	US-07-807-529A-6
57	62	13.8	109	5	PCT-US93-02462-6
58	62	13.8	111	1	US-07-662-193-4
59	62	13.8	111	3	US-08-300-928C-6
60	62	13.8	111	3	US-08-430-944D-6
61	62	13.8	111	3	US-08-430-014-6
62	62	13.8	111	3	US-08-431-184-6
63	62	13.8	145	2	US-08-832-535-2
64	62	13.8	145	3	US-09-019-485-2
65	62	13.8	145	3	US-09-019-485-3
66	62	13.8	145	3	US-09-431-480-9
67	62	13.8	145	3	US-09-617-302-9
68	62	13.8	145	4	US-09-528-436B-2
69	62	13.8	178	2	US-08-791-522-1
70	62	13.8	178	3	US-09-314-777-1
71	61.5	13.7	200	4	US-09-198-452A-571
72	61.5	13.7	214	4	US-09-438-185A-534
73	60.5	13.4	92	1	US-07-807-529A-2
74	60.5	13.4	92	5	PCT-US93-02462-2
75	60.5	13.4	94	1	US-07-662-193-1
76	60.5	13.4	94	3	US-08-300-928C-2
77	60.5	13.4	94	3	US-08-300-928C-11
78	60.5	13.4	94	3	US-08-430-944D-2
79	60.5	13.4	94	3	US-08-430-944D-11
80	60.5	13.4	94	3	US-08-430-014-2
81	60.5	13.4	94	3	US-08-430-014-11
82	60.5	13.4	94	3	US-08-431-184-2
83	60.5	13.4	94	3	US-08-431-184-11
84	60.5	13.4	1441	4	US-09-949-016-10397
85	60	13.3	90	3	US-08-300-928C-14
86	60	13.3	90	3	US-08-430-944D-14
87	60	13.3	90	3	US-08-430-014-14
88	60	13.3	90	3	US-08-431-184-14
89	59	13.1	97	3	US-08-300-928C-10
90	59	13.1	97	3	US-08-430-944D-10
91	59	13.1	97	3	US-08-430-014-10
92	59	13.1	97	3	US-08-431-184-10
93	59	13.1	270	3	US-09-085-305-14
94	58	12.9	92	4	US-09-142-885C-11
95	58	12.9	1023	4	US-09-902-540-12335
96	57.5	12.8	93	4	US-09-198-452A-671
97	57.5	12.8	95	1	US-08-455-896-7
98	57.5	12.8	95	2	US-08-933-149-7
99	57.5	12.8	95	2	US-09-082-343-7
100	57.5	12.8	95	3	US-09-082-253-7

Sequence 1, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 10599, A  
Sequence 21, Appl  
Sequence 26, Appl  
Sequence 16, Appl  
Sequence 21, Appl  
Sequence 14, Appl  
Sequence 32, Appl  
Sequence 19, Appl  
Sequence 24, Appl  
Sequence 23, Appl  
Sequence 18, Appl  
Sequence 22, Appl  
Sequence 17, Appl  
Sequence 5, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 3, Appli  
Sequence 6, Appli  
Sequence 4, Appli  
Sequence 25, Appl  
Sequence 20, Appl  
Sequence 29, Appl  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 4, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 9, Appli  
Sequence 9, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1, Appl  
Sequence 534, App  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 11, Appl  
Sequence 2, Appli  
Sequence 11, Appl  
Sequence 10397, A  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 10, Appl  
Sequence 10, Appl  
Sequence 10, Appl  
Sequence 10, Appl  
Sequence 14, Appl  
Sequence 11, Appl  
Sequence 12335, A  
Sequence 671, App  
Sequence 7, Appli  
Sequence 7, Appli  
Sequence 7, Appli  
Sequence 7, Appli

## ALIGNMENTS

## RESULT 1

US-08-821-451A-4  
; Sequence 4, Application US/08821451A  
; Patent No. 6066724  
; GENERAL INFORMATION:  
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz  
; TITLE OF INVENTION: Human Endometrial Specific Steroid-  
; TITLE OF INVENTION: Binding Factor I, II and III  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/821,451A  
; FILING DATE: March 21, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/014,724  
; FILING DATE: March 21, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULLINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 90 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-821-451A-4

Query Match 100.0%; Score 450; DB 3; Length 90;  
Best Local Similarity 100.0%; Pred. No. 6.6e-49;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDLDPFFISBPLFKLSLAKFDAPPEAVALK 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDLDPFFISBPLFKLSLAKFDAPPEAVALK 60  
QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

## RESULT 2

US-09-263-810-4  
; Sequence 4, Application US/09263810  
; Patent No. 6174992  
; GENERAL INFORMATION:  
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz  
; TITLE OF INVENTION: Human Endometrial Specific Steroid-  
; TITLE OF INVENTION: Binding Factor I, II and III  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/263,810  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/821,451  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULLINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 90 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-09-263-810-4

Query Match 100.0%; Score 450; DB 3; Length 90;  
Best Local Similarity 100.0%; Pred. No. 6.6e-49;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDLDPFFISBPLFKLSLAKFDAPPEAVALK 60  
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELDLDPFFISBPLFKLSLAKFDAPPEAVALK 60  
QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90  
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

## RESULT 3

US-08-912-276-15  
; Sequence 15, Application US/08912276  
; Patent No. 6183952  
; GENERAL INFORMATION:  
; APPLICANT: Billing-Medel, Patricia A.  
; APPLICANT: Cohen, Maurice  
; APPLICANT: Colpitts, Tracey L.  
; APPLICANT: Friedman, Paula N.  
; APPLICANT: Gordon, Julian  
; APPLICANT: Granados, Edward N.  
; APPLICANT: Hodges, Steven C.  
; APPLICANT: Klass, Michael R.  
; APPLICANT: Kratochvil, Jon D.  
; APPLICANT: Roberts-Rapp, Lisa  
; APPLICANT: Russell, John C.  
; APPLICANT: Scroupe, Steven D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE BREAST  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL

```
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSO for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/912,276
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 5972.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6183952e
; US-08-912-276-15

Query Match 100.0%; Score 450; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 6.6e-49;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLSVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60
Db 1 MKLSVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60

Qy 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

RESULT 4
US-09-583-169-4
; Sequence 4, Application US/09583169
; Patent No. 6338948
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CSCCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; APPLICATION NUMBER: US/09/583,169
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,451
; FILING DATE:

; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSO for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/912,276
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-583-169-4

Query Match 100.0%; Score 450; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 6.6e-49;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLSVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60
Db 1 MKLSVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60

Qy 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

RESULT 5
US-09-215-818-6
; Sequence 6, Application US/09215818A
; Patent No. 6379671
; GENERAL INFORMATION:
; APPLICANT: Colpitts, Tracey
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
; TITLE OF INVENTION: DETECTING DISEASES OF THE BREAST
; FILE REFERENCE: 5972.US.P2
; CURRENT APPLICATION NUMBER: US/09/215,818A
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/912,276
; EARLIER FILING DATE: 1997-08-17
; EARLIER APPLICATION NUMBER: 08/697,105
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: 08/912,149
; EARLIER FILING DATE: 1997-08-15
; EARLIER APPLICATION NUMBER: 08/697,106
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-09-215-818-6

Query Match 100.0%; Score 450; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 6.6e-49;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLSVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60
Db 1 MKLSVCLLLVTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPEAAKL 60

Qy 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
Db 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90

RESULT 6
US-09-467-602A-6
; Sequence 6, Application US/09467602A
; Patent No. 6552164
```

```
/ ORGANISM: Homo sapiens
US-09-431-384B-20
Query Match      100.0%; Score 450; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 6.6e-49;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAAKL 60
    |||||||
DB 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAAKL 60
    |||||||
QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
    |||||||
DB 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
    |||||||

RESULT 8
US-08-912-276-23
; Sequence 23, Application US/08912276
; Patent No. 6183952
; GENERAL INFORMATION:
; APPLICANT: Billing-Medel, Patricia A.
; APPLICANT: Cohen, Maurice
; APPLICANT: Colpitts, Tracey L.
; APPLICANT: Friedman, Paula N.
; APPLICANT: Gordon, Julian
; APPLICANT: Granados, Edward N.
; APPLICANT: Hodges, Steven C.
; APPLICANT: Klasse, Michael R.
; APPLICANT: Kratochvil, Jon D.
; APPLICANT: Roberts-Rapp, Lisa
; APPLICANT: Russell, John C.
; APPLICANT: Stroupe, Steven D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/912,276
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 5972.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 69 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6183952e
US-08-912-276-23
Query Match      76.0%; Score 342; DB 3; Length 69;
; TYPE: PRT
```

```
/ ORGANISM: Homo sapiens
US-09-467-602A-6
Query Match      100.0%; Score 450; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 6.6e-49;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAAKL 60
    |||||||
DB 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAAKL 60
    |||||||
QY 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
    |||||||
DB 61 GVKRCTDQMSLQKRSLIAEVLVKILKCSV 90
    |||||||

RESULT 7
US-09-431-384B-20
; Sequence 20, Application US/09431384B
; Patent No. 6770435
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Billing-Medel, Patricia A.
; APPLICANT: Cohen, Maurice
; APPLICANT: Colpitts, Tracey L.
; APPLICANT: Gordon, Julian
; APPLICANT: Granados, Edward N.
; APPLICANT: Hodges, Steven C.
; APPLICANT: Klasse, Michael R.
; APPLICANT: Kratochvil, Jon D.
; APPLICANT: Russell, John C.
; APPLICANT: Scheffel, Christi P.
; APPLICANT: Stroupe, Stephen D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/431,384B
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 5972.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90
; TYPE: PRT
```





```
;
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,451
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-263-810-2

Query Match      61.6%; Score 277; DB 3; Length 90;
Best Local Similarity 58.9%; Pred. No. 3e-27;
Matches 53; Conservative 17; Mismatches 20; Indels 0; Gaps 0;

QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAVAAGL 60
Db 1 MRLSVCLLMVSLALCCYQAHALVCPAVASEITVFLFLSDAAVNLQVAKLNPPPEALAAKL 60

QY 61 GVKRCTDQMSLQKRSIAEVLVKILKCSV 90
Db 61 EVKHCTDQISFKRLSLEKVLVEIVKCGV 90

RESULT 12
US-09-583-169-2
; Sequence 2, Application US/09583169
; Patent No. 6338948
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/583,169
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,451
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-513-999C-4555
; Sequence 4555, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 4555
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -15..-1
; OTHER INFORMATION: score 10.9
; OTHER INFORMATION: seq LSVCLLLVTLALC/CY
US-09-513-999C-4555

Query Match      59.3%; Score 267; DB 4; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.9e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAP 52
Db 1 MKLSVCLLLVTLALCCYQANAEFCPALVSELLDFFFISEPLFKLSLAKFDAP 52

RESULT 14
US-09-471-276-841
; Sequence 841, Application US/09471276
; Patent No. 6822072
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6822072
; FILE REFERENCE: GENSET.025CPI
; CURRENT APPLICATION NUMBER: US/09/471,276
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
```



```

; PRIOR APPLICATION NUMBER: US 08/912,276
; PRIOR FILING DATE: 1997-08-15 DB 4; Length 45;
; PRIOR APPLICATION NUMBER: US 08/697,105
; PRIOR FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: BU101 Synthetic Peptides
US-09-431-384B-27

Query Match 48.7%; Score 219; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.4e-20;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 LAKFDAPPEAAKLGKRCCTDQMSLQKRSLIAEVLVKILKCSV 90
Db 1 LAKFDAPPEAAKLGKRCCTDQMSLQKRSLIAEVLVKILKCSV 45

RESULT 18
US-08-821-451A-25
; Sequence 25, Application US/08821451A
; Patent No. 6066724
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,451A
; FILING DATE: March 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,724
; FILING DATE: March 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-821-451A-25

Query Match 44.9%; Score 202; DB 3; Length 90;
Best Local Similarity 48.4%; Pred. No. 7.4e-18;
Matches 44; Conservative 15; Mismatches 30; Indels 2; Gaps 2;

QY 1 MKLSVCLLLVTLALCCYQANA-EFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAAK 59
Db 1 IELSCLLLIM-LAVCCYEANASQICELVAHETISFLMKSEELKKELEMYNAPPAVEAK 59

RESULT 19
US-08-821-451A-26
; Sequence 26, Application US/08821451A
; Patent No. 6066724
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,451A
; FILING DATE: March 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,724
; FILING DATE: March 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-821-451A-26

Query Match 44.9%; Score 202; DB 3; Length 90;
Best Local Similarity 48.4%; Pred. No. 7.4e-18;
Matches 44; Conservative 15; Mismatches 30; Indels 2; Gaps 2;

QY 1 MKLSVCLLLVTLALCCYQANA-EFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAAK 59
Db 1 IELSCLLLIM-LAVCCYEANASQICELVAHETISFLMKSEELKKELEMYNAPPAVEAK 59

RESULT 20
US-09-263-810-25
; Sequence 25, Application US/09263810
; Patent No. 6174992
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
```

```

; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,810
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,451
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PP257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-263-810-25

```

```

Query Match          44.9%;   Sequence 202;   DB 3; Length 90;
Best Local Similarity 48.4%;   Pred. No. 7.4e-18;
Matches      44; Conservative 15; Mismatches 30; Indels 2; Gaps 2;

QY    1 MKLSVCLLVTLALCCYQANA-ECPALVSELLDFFFIISPLFKLSLAKDPAPPEAAK 59
       :||:|||:: ||:|||:: | : | : | : | : | : | : | : | : | : | : |
Db    1 IELSCLLIM-LAVCCYEANASQICEIVAHETISFLMKSEELKKLEMTNAPPAAVEAK 59
       :||:|||:: ||:|||:: | : | : | : | : | : | : | : | : | : | : |

QY    60 LGVKRCTDOMSLQKRSLIAEVLVKILKKSV 90
       ||||| ||||| ||||| :|||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    60 LEVKRCVDQMNGDRVLVAETLVVFLECGV 90
       ||||| ||||| ||||| :|||: |||: |||: |||: |||: |||: |||: |||:

RESULT 21
US-09-263-810-26
; Sequence 26, Application US/09263810
; Patent No. 6174992
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,810
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821,451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-263-810-26

Query Match          44.9%; Score 202; DB 3; Length 90;
Best Local Similarity 48.4%; Pred. No. 7.4e-18;
Matches 44; Conservative 15; Mismatches 30; Indels 2; Gaps 2;

QY      1 MKLSVCLLLVTALCCYQANA-BFCPALVSELLDFFFISEPLFKLSIAKFDAPPEAAVK 59
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      1 IEISLCLLIIM-LAVCCYEANASQICELVAHETISFLMKSEBELKKELEMYNAPPAAVEAK 59
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      60 LGVKRCTDQMSLOQRSLIAEVLVKILKKCSV 90
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      60 LEVKRCVDQMSGNDRLVVAETLVVFLECGV 90
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 22
US-09-583-169-25
Sequence 25, Application US/09583169
Patent No. 6338948
GENERAL INFORMATION:
APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
BIOLOGICAL FUNCTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/583,169
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821,451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 25:

```



ADDRESSEE: HOWELL & HAFERKAMP, L.C.  
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400  
CITY: ST. LOUIS  
STATE: MISSOURI  
COUNTRY: USA  
ZIP: 63105-1817  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,149  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: HENDERSON, MELODIE W.  
REGISTRATION NUMBER: 37,848  
REFERENCE/DOCKET NUMBER: 6029-6040  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314) 727-5188  
TELEFAX: (314) 727-6092  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 91 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-933-149-8

Query Match 24.4%; Score 110; DB 2; Length 91;  
Best Local Similarity 31.0%; Pred. No. 2.5e-06;  
Matches 26; Conservative 17; Mismatches 41; Indels 0; Gaps 0;  
QY 1 MKLSVCILLVTLALCCYQANAEFCALVSELDDFFFISEPLFKLSLAKEDAPPEAVAACL 60  
Db 1 MKLAVTLVTLALCCSSASAEICPSQQRVIETLLMDTPSSYEAAMELFSPDQDMREAGA 60  
QY 61 GVKRCDQMSLQKRSLIAEVLVKI 84  
Db 61 QLKXLDVTLPOKPRESLIKLMEKI 84

Search completed: September 26, 2005, 08:32:28  
Job time : 26.0549 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 07:44:37 ; Search time 92.967 Seconds  
(without alignments)  
495.736 Million cell updates/sec

Title: US-09-975-502A-6

Perfect score: 450

Sequence: 1 MKLSVCLLVLTALCCVQAN.....LQKSLIAEVLVKILKCSV 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	450	100.0	90	1	LPPB_HUMAN
2	267	59.3	90	1	LPPA_HUMAN
3	262	58.2	102	2	Q7PCX8
4	249	55.3	90	2	Q9GK67
5	245	54.4	96	2	Q8WMS2
6	238	52.9	83	1	LPPD_HUMAN
7	231.5	51.4	91	2	Q9GK66
8	229	50.9	90	2	Q7YS38
9	226	50.2	90	2	Q9GK65
10	209.5	46.6	111	2	Q863D3
11	206	45.8	111	1	PSC1_MOUSE
12	206	45.8	111	1	PSC1_RAT
13	171	38.0	112	1	PSC2_RAT
14	158	35.1	108	2	Q7M743
15	110	24.4	91	1	UTER_HUMAN
16	102	22.7	91	1	UTER_RABIT
17	101	22.4	96	1	UTER_MOUSE
18	100	22.2	91	1	UTER_LEPCA
19	100	22.2	96	1	UTER_RAT
20	91	20.2	95	2	Q8TD33
21	90	20.0	91	2	Q8MKG2
22	85	18.9	96	2	Q8VD96
23	80.5	17.9	139	1	UGR1_MOUSE
24	78.5	17.4	317	2	Q8VFH6
25	77.5	17.2	94	2	Q85702
26	75.5	16.8	93	1	UGR1_HUMAN
27	74	16.4	94	2	Q7M742
28	72	16.0	362	2	Q7YU06
29	70.5	15.7	93	2	Q65C83
30	70.5	15.7	320	2	Q8VF26
31	70	15.6	173	2	Q7Xtk2

32	70	15.6	405	2	Q88TS3	Q88ts3 lactobacill
33	70	15.6	486	2	Q84MF5	Q84mf5 oryza sativ
34	70	15.6	1495	2	Q6ATE0	Q6ate0 oryza sativ
35	70	15.6	1501	2	Q6AT86	Q6at86 oryza sativ
36	70	15.6	1501	2	Q7XQE2	Q7xqe2 oryza sativ
37	70	15.6	1516	2	Q7X798	Q7x798 oryza sativ
38	69	15.3	144	1	CYTF_MOUSE	Q89098 mus musculu
39	69	15.3	167	2	Q9OWL5	Q9owl5 mus musculu
40	69	15.3	392	2	Q8M78	Q8m78 pseudomonas
41	69	15.3	1622	2	Q48908	Q48908 arabidopsis
42	69	15.3	1623	2	Q48907	Q48907 arabidopsis
43	69	15.3	1623	2	Q64590	Q64590 arabidopsis
44	68.5	15.2	1078	2	Q7RCN2	Q7rcn2 plasmodium
45	68.5	15.2	1623	2	Q22449	Q22449 arabidopsis
46	68	15.1	290	2	Q8W3C2	Q8w3c2 oryza sativ
47	68	15.1	290	2	Q7XHG7	Q7xhg7 oryza sativ
48	68	15.1	365	2	Q7YUQ7	Q7yuq7 trypanosoma
49	68	15.1	365	2	Q7YU08	Q7yu08 trypanosoma
50	68	15.1	367	2	Q7YU08	Q7yu08 trypanosoma
51	68	15.1	369	2	Q7YUQ1	Q7yuq1 trypanosoma
52	68	15.1	369	2	Q7YUQ2	Q7yuq2 trypanosoma
53	68	15.1	369	2	Q7YUQ4	Q7yuq4 trypanosoma
54	68	15.1	1369	2	Q65WU1	Q65wu1 oryza sativ
55	67.5	15.0	194	2	Q6QZV3	Q6qzv3 ornithodor
56	67.5	15.0	338	2	Q8CDW5	Q8cdw5 m mus muscu
57	67.5	15.0	347	2	Q8CCH2	Q8cch2 m mus muscu
58	67	14.9	74	2	Q8W2V1	Q8w2v1 oryza sativ
59	67	14.9	74	2	Q7G730	Q7g730 oryza sativ
60	67	14.9	182	2	Q6UU90	Q6uu90 oryza sativ
61	67	14.9	1516	2	Q7XQU9	Q7xqu9 oryza sativ
62	66.5	14.8	434	2	Q9VZA6	Q9vza6 drosophila
63	66.5	14.8	578	2	Q9XVG0	Q9xvg0 caenorhabdi
64	66.5	14.8	703	2	Q7S6S1	Q7s6s1 neurospora
65	66.5	14.8	3079	1	IRA2_YEAST	P19158 saccharomyc
66	66	14.7	102	2	Q8USC3	Q8usc3 agrobacteri
67	66	14.7	351	2	Q7YU04	Q7yu04 trypanosoma
68	66	14.7	474	1	GATB_WOLPM	P61349 wolbachia p
69	66	14.7	1040	2	Q6CTA3	Q6cta3 kluyveromyc
70	66	14.7	1516	2	Q6FQ06	Q6fq06 oryza sativ
71	65.5	14.6	310	2	Q96649	Q96649 trypanosoma
72	65.5	14.6	310	2	Q97466	Q97466 trypanosoma
73	65.5	14.6	364	2	Q7YUQ5	Q7yuq5 trypanosoma
74	65.5	14.6	415	1	AL7B_HUMAN	Q9v614 homo sapien
75	65.5	14.6	415	1	AL7B_MACFA	Q95jk8 macaca fasc
76	65	14.4	470	2	Q8S9A4	Q8s9a4 phaseolus a
77	64.5	14.3	93	2	Q71MM7	Q71mm7 rattus norv
78	64.5	14.3	369	2	Q7YUQ3	Q7yuq3 trypanosoma
79	64.5	14.3	465	2	Q6KZM4	Q6kzm4 picophilus
80	64.5	14.3	766	2	Q6PB57	Q6pb57 mus musculu
81	64.5	14.3	1081	2	Q8IKS3	Q8iks3 plasmodium
82	64.5	14.3	2664	1	ANRY_HUMAN	Q6ub99 homo sapien
83	64	14.2	213	2	Q9VBQ4	Q9vbq4 drosophila
84	64	14.2	366	2	Q9U0W2	Q9u0w2 leishmania
85	64	14.2	787	1	ADR1_ARATH	Q9fw44 arabidopsis
86	64	14.2	1038	2	Q6ATC8	Q6atc8 oryza sativ
87	64	14.2	2634	2	Q6BG52	Q6bg52 paramacium
88	63.5	14.1	1718	2	Q6CAB3	Q6cab3 yarrowia li
89	63	14.0	115	2	Q9D4K3	Q9d4k3 mus musculu
90	63	14.0	256	1	Y079_PHOLL	Q7n475 photorhabdu
91	63	14.0	436	2	Q87PB4	Q87pb4 vibrio para
92	63	14.0	442	1	TIG_BUCAP	Q8k991 buchera ap
93	63	14.0	446	2	Q9NVH3	Q9nvh3 homo sapien
94	63	14.0	467	2	Q8J1Q1	Q8j1q1 seriola qui
95	63	14.0	473	2	Q9H384	Q9h384 homo sapien
96	63	14.0	505	2	Q96EF4	Q96ef4 homo sapien
97	63	14.0	547	2	Q816X1	Q816x1 ancylostoma
98	63	14.0	676	1	PK26_HUMAN	Q8bq70 homo sapien
99	63	14.0	735	1	Q9H7V5	Q9h7v5 homo sapien
100	63	14.0	756	2	Q96ST4	Q96st4 homo sapien

ALIGNMENTS

```

RESULT 1
LPPB_HUMAN
ID LPPB_HUMAN STANDARD; PRT; 90 AA.
AC O95969;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Lipophilin B precursor (Secretoglobin family 1D member 2).
GN Name=SCGB1D2; Synonyms=LIPHB;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99167354; PubMed=10066439; DOI=10.1006/bbrc.1999.0274;
RA Zhao C., Nguyen T., Yusifov T., Glasgow B.J., Lehrer R.I.;
RT "Lipophilins: human peptides homologous to rat prostatein.";
RL Biochem. Biophys. Res. Commun. 256:147-155(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: May bind androgens and other steroids, may also bind
CC estramustine, a chemotherapeutic agent used for prostate cancer.
CC May be under transcriptional regulation of steroid hormones.
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC muscle. Expressed as well in thymus, trachea, kidney, steroid
CC responsive tissues (prostate, testis, uterus, breast and ovary)
CC and salivary gland.
CC -1- SIMILARITY: Belongs to the uteroglobin family. Lipophilin
CC subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ224172; CAA11864.1; -.
DR ENBL; BC069290; AAH69290.1; -.
DR GenBank; HGNC:18396; SCGB1D2.
DR InterPro; IPR006039; Utergl.
DR InterPro; IPR000329; Uteroglobulin subf.
DR InterPro; IPR006038; Uteroglobulin_supf.
DR Pfam; PF01099; Uterogloblin; 1.
DR PRINTS; PR00486; UTEROGLOBIN.
DR SMART; SM00096; UTG; 1.
KW Signal.

FT SIGNAL 1 21 Potential.
FT CHAIN 22 90 Lipophilin B.
FT SEQUENCE 90 AA; 9925 MW; 17BB555ED035D1AF CRC64;
Query Match 100.0%; Score 450; DB 1; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.4e-43;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLSVCLLIATVTLALCCYQANAEFCPALVSELDFFIISPLFKLSLAKFDAPPEAAAKL 60
DB 1 MKLSVCLLIATVTLALCCYQANAEFCPALVSELDFFIISPLFKLSLAKFDAPPEAAAKL 60
QY 61 GVKRCTDQMSLQKRSIIAEVLVKILKCSV 90
DB 61 GVKRCTDQMSLQKRSIIAEVLVKILKCSV 90

RESULT 2
LPPA_HUMAN
ID LPPA_HUMAN STANDARD; PRT; 90 AA.
AC O95968;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Lipophilin A precursor (Secretoglobin family 1D member 1).
GN Name=SCGB1D1; Synonyms=LIPHA;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99167354; PubMed=10066439; DOI=10.1006/bbrc.1999.0274;
RA Zhao C., Nguyen T., Yusifov T., Glasgow B.J., Lehrer R.I.;
RT "Lipophilins: human peptides homologous to rat prostatein.";
RL Biochem. Biophys. Res. Commun. 256:147-155(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 22-90, AND MASS SPECTROMETRY.
RX MEDLINE=98385871; PubMed=9720917; DOI=10.1016/S0014-5793(98)00852-7;
RA Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Qu X.-D., Martin D.,
RA Glasgow B.J.;
RT "Lipophilin, a novel heterodimeric protein of human tears.";
RL FEBS Lett. 432:163-167(1998).
CC -1- FUNCTION: May bind androgens and other steroids, may also bind
CC estramustine, a chemotherapeutic agent used for prostate cancer.
CC May be under transcriptional regulation of steroid hormones.
CC -1- SUBUNIT: Heterodimer of a lipophilin A and a lipophilin C
CC (mammoglobin B) monomer associated head to head.
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC -1- TISSUE SPECIFICITY: Expressed in lachrymal gland, thymus, kidney,

```

CC testis, ovary and salivary gland.  
 CC -1- MASS SPECTROMETRY: MW=7574.63; METHOD=Electrospray; RANGE=22-90;  
 CC NOTE=Ref.3.  
 CC -1- SIMILARITY: Belongs to the uteroglobin family. Lipophilin  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AJ224171; CA11863.1; -.  
 CC EMBL; BC069170; AAH69170.1; -.  
 CC EMBL; BC069289; AAH69289.1; -.  
 CC Genew; HGNC:18395; SCGB1D1.  
 CC GO; GO:0005615; C:extracellular space; TAS.  
 CC InterPro; IPR006039; Utergl.  
 CC InterPro; IPR000329; Uteroglobin subf.  
 CC InterPro; IPR006038; Uteroglobin\_supf.  
 CC Pfam; PF01099; Uteroglobin.1.  
 CC PRINTS; PR00486; UTEROGLGBIN.  
 CC SMART; SM00096; UTG; 1.  
 CC KW Direct protein sequencing; Signal.  
 CC FT SIGNAL 1 21 Lipophilin A.  
 CC FT CHAIN 22 90  
 CC SQ SEQUENCE 90 AA; 9898 MW; DF2D4F7565A87D34 CRC64;  
 CC  
 CC Query Match 59.3%; Score 267; DB 1; Length 90;  
 CC Best Local Similarity 60.2%; Pred. No. 1.6e-22;  
 CC Matches 53; Conservative 12; Mismatches 23; Indels 0; Gaps 0;  
 CC  
 CC 1 MKLSVCLLLVTALCCYQNAEFCPALVSELLDFFIFSEPLFKLSLAKFDAPPEAVAAL 60  
 CC 1 MRLSVTALLVTALCYEANAIVCPFTAADLTFFYFDLLYRLSLAKYNAPPEAVAAM 60  
 CC  
 CC 61 GVKRCTDMSLQKRSLIAEVLVKILKCSV 88  
 CC 61 EVKCVDTMAVEKRVLTITKLGIAEKC 88  
 CC  
 CC RESULT 3  
 CC Q7PCK8 PRELIMINARY; PRT; 102 AA.  
 CC ID Q7PCK8  
 CC AC Q7PCK8;  
 CC DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 CC DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 CC DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 CC DE LppAB.  
 CC OS Bos taurus (Bovine).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC OC Bovinae; Bos.  
 CC OX NCBI\_TaxID=9913;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=22338213; PubMed=12406855;  
 CC RA Reynolds S.D., Reynolds P.R., Pryhuber G.S., Finder J.D., Stripp B.R.;  
 CC RT "Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in  
 CC RL mouse and human airways."  
 CC RL Am. J. Respir. Crit. Care Med. 166:1498-1509(2002).  
 CC CC -1- MISCCELLANEOUS: The sequence shown here is derived from an  
 CC EMBL/GenBank/DDAJ third party annotation (TPA) entry.  
 CC CC EMBL; BK000199; DAA00356.1; -.  
 CC DR InterPro; IPR006038; Uteroglobin\_supf.  
 CC DR Pfam; PF01099; Uteroglobin.1.  
 CC SQ SEQUENCE 102 AA; 11294 MW; 422D4BE83D1018A5 CRC64;  
 CC  
 CC Query Match 58.2%; Score 262; DB 2; Length 102;  
 CC Best Local Similarity 55.6%; Pred. No. 6.9e-22;  
 CC Matches 50; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

Qy 1 MKLSVCLLLVTALCCYQNAEFCPALVSELLDFFIFSEPLFKLSLAKFDAPPEAVAAL 60  
 Db 1 MRLSVTALLVTALCYEANAIVCPFTAADLTFFYFDLLYRLSLAKYNAPPEAVAAM 60  
 Qy 61 GVKRCTDMSLQKRSLIAEVLVKILKCSV 90  
 Db 61 EVKCVDTMAVEKRVLTITKLGILNCTV 90  
 RESULT 4  
 Q9GK67 PRELIMINARY; PRT; 90 AA.  
 ID Q9GK67  
 AC Q9GK67;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Lipophilin AL.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lacimal gland;  
 RA Zhao C., Nguyen T.X., Lehrer R.I.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF308614; AAG42802.1; -.  
 DR InterPro; IPR006039; Utergl.  
 DR InterPro; IPR006038; Uteroglobin\_supf.  
 DR Pfam; PF01099; Uteroglobin.1.  
 DR SMART; SM00096; UTG; 1.  
 DR SQ SEQUENCE 90 AA; 9762 MW; 6D34F9540C1FF742 CRC64;  
 CC  
 CC Query Match 55.3%; Score 249; DB 2; Length 90;  
 CC Best Local Similarity 56.7%; Pred. No. 1.8e-20;  
 CC Matches 51; Conservative 15; Mismatches 24; Indels 0; Gaps 0;  
 CC  
 CC 1 MKLSVCLLLVTALCCYQNAEFCPALVSELLDFFIFSEPLFKLSLAKFDAPPEAVAAL 60  
 CC 1 MKLLVPLLVALALGCVYEAADAAACPAFLDSVGLFDPKFPYRQLAKYDAPPEAVEAL 60  
 CC  
 CC 61 GVKRCTDMSLQKRSLIAEVLVKILKCSV 90  
 CC 61 QVKECTDEIDKGRVLAALVTIKIVRECA 90  
 CC  
 CC RESULT 5  
 CC Q8WMS2 PRELIMINARY; PRT; 96 AA.  
 ID Q8WMS2  
 AC Q8WMS2;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Lipophilin AL2.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lacimal gland;  
 RA Lind-Ayres M.R., Crow J.M., Steelman C.A., Nelson J.D.,  
 RA Remington S.G.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY063770; AAL40859.1; -.  
 DR Pfam; PF01099; Uteroglobin.1.  
 DR PRINTS; PR00486; UTEROGLGBIN.  
 DR SMART; SM00096; UTG; 1.  
 DR SQ SEQUENCE 96 AA; 10510 MW; 48EC2972D0EA78C4 CRC64;  
 CC  
 CC Query Match 54.4%; Score 245; DB 2; Length 96;  
 CC Best Local Similarity 52.8%; Pred. No. 5.6e-20;



```
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Lacrimal gland;
RC Remington S.C., Nelson J.D.;
RA Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY303698; AAP59424.1; -.
DR InterPro; IPR006038; Uterogloblin_supf.
DR Pfam; PF01099; Uterogloblin; 1.
SQ SEQUENCE 90 AA; 9957 MW; 4118D604035EB1BC CRC64;

Query Match 50.2%; Score 226; DB 2; Length 90;
Best Local Similarity 51.1%; Pred. No. 7.6e-18;
Matches 46; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MKLSVCLLVLTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPAAVAKL 60
Db 1 MRLSVSLMTVTALCCYEGNALVCPALLAENFGYLFNKDVFRLQLAKFMPPREAAEALL 60
Qy 61 GVKRCTDMSLQKRSLIAEVLVKILKCSV 90
Db 61 TVKCKTDGMPFEKRNLIAGALGEVVLQCPV 90

RESULT 9
O9GK65 PRELIMINARY; PRT; 90 AA.
AC O9GK65
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Lipophilin BL
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lacrimal gland;
RA Zhao C., Nguyen T.X., Lehrer R.I.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF308616; AAG42804.1; -.
DR InterPro; IPR000329; Uterogloblin_subf.
DR InterPro; IPR006038; Uterogloblin_supf.
DR Pfam; PF01099; Uterogloblin; 1.
DR PRINTS; PR00486; UTEROGLOBIN.
SQ SEQUENCE 90 AA; 9943 MW; 410900DF7F3EB1BC CRC64;

Query Match 50.2%; Score 226; DB 2; Length 90;
Best Local Similarity 51.1%; Pred. No. 7.6e-18;
Matches 46; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MKLSVCLLVLTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPAAVAKL 60
Db 1 MRLSVSLMTVTALCCYEGNALVCPALLAENFGYLFNKDVFRLQLAKFMPPREAAEALL 60
Qy 61 GVKRCTDMSLQKRSLIAEVLVKILKCSV 90
Db 61 TVKCKTDGMPFEKRNLIAGALGEVVLQCPV 90

RESULT 10
O863D3 PRELIMINARY; PRT; 111 AA.
AC O863D3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pheromaxein A subunit precursor.
GN Name=PHEROA;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
```

```
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Submaxillary salivary glands;
RA Austin C.J., Emberson L., Nicholls P.;
RT "Purification and characterisation of pheromaxein, the porcine
steroid-binding protein.";
RL Eur. J. Biochem. 271:2593-2606 (2004).
DR EMBL; AJ537467; CAD60373.1; -.
DR InterPro; IPR006038; Uterogloblin_supf.
DR Pfam; PF01099; Uterogloblin; 1.
KW Signal.
FT SIGNAL
FT CHAIN 22 111 pheromaxein A subunit.
SQ SEQUENCE 111 AA; 12351 MW; 4106A01902194EC9 CRC64;

Query Match 46.6%; Score 209.5; DB 2; Length 111;
Best Local Similarity 49.4%; Pred. No. 7.1e-16;
Matches 44; Conservative 18; Mismatches 26; Indels 1; Gaps 1;

Qy 1 MKLSVCLLVLTALCCYQANAEFCPALVSELDFFFISEPLFKLSLAKFDAPPAAVAKL 60
Db 1 MRLSLTVLLVTALCCYEAHGIVCRAIVKFSAPLWKPDIEIYKPELELFGAPPEAVDAKM 60
Qy 61 GVKRCTDMSLQKRSLIAEVLVKIL-KKC 88
Db 61 KVKQCANGISFKKILLTKTLVELLVKCC 89

RESULT 11
PSC1 MOUSE STANDARD; PRT; 111 AA.
ID PSC1_MOUSE
AC P08089; 2004 (Rel. 44, Created)
DT 03-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Prostatic steroid-binding protein C1 chain precursor (Prostatein
peptide C1).
DE Name=Psbpc1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Kaushal V., Chatta G.S.;
RT "Sequence of C1 chain of mouse prostate steroid binding protein.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Part of prostatein which is the major secretory
glycoprotein of ventral prostate gland (By similarity).
CC -1- SUBUNIT: Prostatein is composed of three different peptides called
C1, C2 and C3 (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- SIMILARITY: Belongs to the uterogloblin family. Lipophilin
subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
CC EMBL; AF277385; AAG17693.1; -.
DR MGD; MGI:2150387; Psbpc1.
DR PROSITE; PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE; PS00404; UTEROGLOBIN_2; FALSE_NEG.
KW Signal; Steroid-binding.
FT SIGNAL 1 23 By similarity.
FT CHAIN 24 111 Prostatic steroid-binding protein C1
chain.
SQ SEQUENCE 111 AA; 12763 MW; 2C1D11D003952945 CRC64;
```

```

Query Match          45.8%; Score 206; DB 1; Length 111;
Best Local Similarity 49.5%; Pred. No. 1.8e-15;
Matches 45; Conservative 14; Mismatches 30; Indels 2; Gaps 2;

QY 1 MKLSVCLLVTLALCCYQANA-EFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAVAK 59
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 4 IKLSCLLLIM-LAVCCYEANASQICELVAHETISFLMKSEELKKELEMYNAPPAVAK 62
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 60 LGVKRCTDOMSLQKRSLIAEVLVKILKCSV 90
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 63 LEVKRCVDQMSNGDLVVAETLVYIFLECGV 93
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 12
PSC1 RAT
ID _PSC1 RAT STANDARD; PRT; 111 AA.
AC P02782; Q63469;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Prostatic steroid-binding protein C1 chain precursor (Prostatein
DE peptide C1).
DE Name=Psbpc1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82220075; PubMed=6896362;
RA Parker M.G., Needham M., White R.;
RT "Prostatic steroid binding protein: gene duplication and steroid
RT binding.";
RL Nature 298:92-94(1982).
RN [2]
RP REVISIONS.
RA Parker M.G.;
RL Submitted (JUL-1983) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=83234456; PubMed=6688048;
RA Delaey B., Dirckx L., Peeters B., Volckaert G., Mous J., Heyns W.,
RA Rombaerts W.;
RT "The nucleotide sequence of cDNA complementary to the C1 component of
RT rat prostatic binding protein.";
RL Eur. J. Biochem. 133:645-649(1983).
RN [4]
RP SEQUENCE OF 24-111.
RX MEDLINE=82164744; PubMed=7200013;
RA Peeters B., Heyns W., Mous J., Rombaerts W.;
RT "Structural studies on rat prostatic binding protein. The primary
RT structure of component C1 from subunit F.";
RL Eur. J. Biochem. 123:55-62(1982).
RN [5]
RP SEQUENCE OF 24-111.
RX MEDLINE=82075873; PubMed=7198120;
RA Liao S., Chen C., Huang I.-Y.;
RT "Prostate alpha-protein. Complete amino acid sequence of the component
RT that inhibits nuclear retention of the androgen-receptor complex.";
RL J. Biol. Chem. 257:122-125(1982).
RN [6]
RP SEQUENCE OF 13-65 FROM N.A.
RA Delaey B., Rombaerts W., Volckaert G., Peeters B., Mous J., Heyns W.;
RT "Identification of a complementary-DNA clone containing part of the
RT sequence information for the C-1-polypeptide of rat prostatic binding
RT protein.";
RL Biochem. Soc. Trans. 10:51-51(1982).
CC CC -1- FUNCTION: Part of prostatein which is the major secretory
CC glycoprotein of ventral prostate gland.
CC CC -1- SUBUNIT: Prostatein is composed of three different peptides called
CC C1, C2 and C3. These form covalent C1:C3 (F) and C2:C3 (S)
CC heterodimers whose noncovalent association forms tetrameric

```

```

(C1:C3/C2:C2) prostatein molecules.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: The heterodimer can bind non-polar steroids,
CC cholesterol and a group of small proline-rich peptides.
CC -1- SIMILARITY: Belongs to the uteroglobin family. Lipophilin
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01255; CAA24568.1; -
DR EMBL; J00774; AAA41969.1; -
DR EMBL; J00773; AAA41969.1; JOINED.
DR EMBL; V01545; CAA24787.1; -
DR PIR; A93286; BORT1.
DR InterPro; IPR006039; Utergl.
DR InterPro; IPR000329; Uteroglobin_subf.
DR InterPro; IPR006038; Uteroglobin_supf.
DR Pfam; PF01099; Uteroglobin; 1.
DR SMART; SM00096; UTG; 1.
DR PROSITE; PS00403; UTEROGLIBIN_1; FALSE NEG.
DR PROSITE; PS00404; UTEROGLIBIN_2; FALSE NEG.
KW Direct protein sequencing; Signal; Steroid-binding.
FT SIGNAL 1 23
FT CHAIN 24 111
FT PROSTATEIN steroid-binding protein C1
FT chain.
FT CONFLICT 5 5 K -> E (in Ref. 1; AAA41969).
FT CONFLICT 15 15 A -> S (in Ref. 6).
FT CONFLICT 17 18 CC -> GG (in Ref. 6).
FT CONFLICT 74 74 N -> D (in Ref. 5).
SQ SEQUENCE 111 AA; 12763 MW; 2C1D1D003952945 CRC64;

Query Match          45.8%; Score 206; DB 1; Length 111;
Best Local Similarity 49.5%; Pred. No. 1.8e-15;
Matches 45; Conservative 14; Mismatches 30; Indels 2; Gaps 2;

QY 1 MKLSVCLLVTLALCCYQANA-EFCPALVSELDFFIFSEPLFKLSLAKFDAPPEAAVAK 59
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 4 IKLSCLLLIM-LAVCCYEANASQICELVAHETISFLMKSEELKKELEMYNAPPAVAK 62
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 60 LGVKRCTDOMSLQKRSLIAEVLVKILKCSV 90
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 63 LEVKRCVDQMSNGDLVVAETLVYIFLECGV 93
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 13
PSC2 RAT
ID _PSC2 RAT STANDARD; PRT; 112 AA.
AC P02781;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Prostatic steroid-binding protein C2 chain precursor (Prostatein
DE peptide C2).
DE Name=Psbpc2;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87146484; PubMed=2881277;
RA Delaey B., Dirckx L., Decourt J.-L., Claessens F., Peeters B.,
RA Rombaerts W.;
RT "Rat prostatic binding protein: the complete sequence of the C2 gene
RT and its flanking regions.";
RL Nucleic Acids Res. 15:1627-1641(1987).
RN [2]

```

RP SEQUENCE OF 21-112.  
RX MEDLINE=83209613; PubMed=6343081;  
RA Peeters B., Heyns W., Mous J., Rombauts W.;  
RT "Structural studies on rat prostatic binding protein. The primary  
RL structure of component C2 from subunit S.";  
RN Eur. J. Biochem. 132:669-679(1983).  
[3]  
RP SEQUENCE OF 1-100 FROM N.A.  
RX MEDLINE=82220075; PubMed=6896362;  
RA Parker M., Needham M., White R.;  
RT "Prostatic steroid binding protein: gene duplication and steroid  
RT binding.";  
RL Nature 298:92-94(1982).  
CC -1- FUNCTION: Part of prostatein which is the major secretory  
CC glycoprotein of ventral prostate gland.  
CC -1- SUBUNIT: Prostatein is composed of three different peptides called  
CC C1, C2 and C3. These form covalent C1:C3 (F) and C2:C3 (S)  
CC heterodimers whose noncovalent association forms tetrameric  
CC (C1:C3/C3:C2) prostatein molecules.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- PM: Linked by three disulfide bonds to C3.  
CC -1- PM: The N-terminus is blocked.  
CC -1- MISCELLANEOUS: The heterodimer can bind non-polar steroids,  
CC cholesterol and a group of small proline-rich peptides.  
CC -1- SIMILARITY: Belongs to the uteroglobin family. Lipophilin  
CC subfamily.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; X05034; CAA28708.1; -;  
DR EMBL; V01256; CAA24569.1; -;  
DR EMBL; J00776; AAS1641.1; -;  
DR PIR; A03251; BORT2.  
DR InterPro; IPR000329; Uteroglobin subf.  
DR InterPro; IPR006038; Uteroglobin supf.  
DR PROSITE; PS00403; UTEROGLIBIN\_1; FALSE\_NEG.  
DR PROSITE; PS00404; UTEROGLIBIN\_2; FALSE\_NEG.  
KW Direct protein sequencing; Pyridone carboxylic acid; Signal;  
FT Steroid-binding.  
FT SIGNAL 1 20  
FT CHAIN 21 112 Prostatic steroid-binding protein C2  
FT chain.  
FT MOD\_RES 21 21 Pyridone carboxylic acid (Probable).  
FT DISULFID 28 28 Interchain (with C3) (Probable).  
FT DISULFID 69 69 Interchain (with C3) (Probable).  
FT DISULFID 92 92 Interchain (with C3) (Probable).  
FT CONFLICT 26 26 Missing (in Ref. 3).  
FT CONFLICT 88 88 I -> T (in Ref. 3).  
FT CONFLICT 96 112 VWLQINPRGRWFSEIN -> YGYK (in Ref. 3).  
SQ SEQUENCE 112 AA; 12828 MW; DA65A6A82E677864 CRC64;  
Query Match 38.0%; Score 171; DB 1; Length 112;  
Best Local Similarity 42.1%; Pred. No. 1.7e-11;  
Matches 40; Conservative 13; Mismatches 36; Indels 6; Gaps 2;  
Qy 1 MKLSVCLLLVTALCCYQANA-----AFCPALVSELDFFPISPLPKLSLAKFDAPPEAA 55  
Db 1 MRLSLCLLTI-LVCCYEANGQTLAGQVCQALQDVITFLINPEELKRELEEFDPPEA 59  
Qy 56 VAAKLGKRCCTDQMSLQKRSIAEVLVKILKCSV 90  
Db 60 VEANLKVRCINKIMYGDRLSMGTSLVFIMLKCDV 94  
RESULT 14  
Q7M743 PRELIMINARY; PRT; 108 AA.  
ID Q7M743

AC Q7M743;  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE PBPCIBS.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22338213; PubMed=12406855;  
RA Reynolds S.D., Reynolds P.R., Pyhuber G.S., Finder J.D., Stripp B.R.;  
RT "Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in  
RT mouse and human airways.";  
RL Am. J. Respir. Crit. Care Med. 166:1498-1509(2002).  
CC -1- MISCELLANEOUS: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ third party annotation (TPA) entry.  
DR EMBL; BK000198; DAA00355.1; -;  
DR InterPro; IPR006038; Uteroglobin\_supf.  
DR Pfam; PF01039; Uteroglobin; 1. CDC594842BEE284D CRC64;  
SQ SEQUENCE 108 AA; 12410 MW; CDC594842BEE284D CRC64;  
Query Match 35.1%; Score 158; DB 2; Length 108;  
Best Local Similarity 41.6%; Pred. No. 5.1e-10;  
Matches 37; Conservative 14; Mismatches 36; Indels 2; Gaps 2;  
Qy 1 MKLSVCLLLVTALCCYQANA-EFCPALVSELDFFPISPLPKLSLAKFDAPPEAAK 59  
Db 4 VRLSPCLLTI-LVCCYETNAGKICDAFWSESLRSEDLKKELEYKSAKPKAVEAK 62  
Qy 60 LGVKRCCTDQMSLQKRSIAEVLVKILKCC 88  
Db 63 LEVKQCVDQMHYLDLRERVLALMYVSLRC 91  
RESULT 15  
UTER\_HUMAN STANDARD; PRT; 91 AA.  
ID UTER\_HUMAN  
AC P11684; Q9UCM2; Q9UCM4;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Clara cell phospholipid-binding protein precursor (CCPB) (Clara cells  
DE 10 kDa secretory protein) (CC10) (Uteroglobin) (Urine protein 1)  
DE (UPI).  
GN Name=SCGB1A1; Synonyms=CC10, CCSP, UGB;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=89000784; PubMed=3167058; DOI=10.1016/0167-4781(88)90129-7;  
RA Singh G., Katyal S.L., Brown W.E., Phillips S., Kennedy A.L.,  
RA Anthony J., Squeglia N.;  
RT "Amino-acid and cDNA nucleotide sequences of human Clara cell 10 kDa  
RT protein.";  
RL Biochim. Biophys. Acta 950:329-337(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=95250987; PubMed=7733299;  
RA Hay J.G., Danel C., Chu C., Crystal R.G.;  
RT "Human CC10 gene expression in airway epithelium and subchromosomal  
RT locus suggest linkage to airway disease.";  
RL Am. J. Physiol. 268:L565-L575(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,



RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RN SEQUENCE OF 1-18 FROM N.A.  
 RP MEDLINE=93250776; PubMed=1284526;  
 RA Wolf M., Klug J., Hackenberg R., Gessler M., Grzeschik K.-H.,  
 RA Beato M., Suske G.;  
 RT "Human CC10, the homologue of rabbit uteroglobin: genomic cloning,  
 RT chromosomal localization and expression in endometrial cell lines.";  
 RL Hum. Mol. Genet. 1:371-378(1992).  
 RN [5]  
 RN SEQUENCE OF 22-74.  
 RP TISSUE=Urine;  
 RX MEDLINE=93016476; PubMed=1400743;  
 RA Okutani R., Itoh Y., Hirata H., Kasahara T., Mukaide N., Kawai T.;  
 RT "Simple and high-yield purification of urine protein 1 using  
 RT immunoaffinity chromatography: evidence for the identity of urine  
 RT protein 1 and human Clara cell 10-kilodalton protein.";  
 RL J. Chromatogr. A 577:25-35(1992).  
 RN [6]  
 RN SEQUENCE OF 22-45.  
 RP TISSUE=Urine;  
 RX MEDLINE=93009001; PubMed=1395029; DOI=10.1016/0009-8981(92)90122-7;  
 RA Bernard A., Roels H., Lauwerys R., Witters R., Gielens C.,  
 RA Soumillion A., Van Damme J., De Ley M.;  
 RT "Human urinary protein 1: evidence for identity with the Clara cell  
 RT protein and occurrence in respiratory tract and urogenital  
 RT secretions.";  
 RL Clin. Chim. Acta 207:239-249(1992).  
 RN [7]  
 RN SEQUENCE OF 22-33.  
 RP MEDLINE=21648993; PubMed=11788998;  
 RX DOI=10.1002/1615-9861(200201)2:1<112::AID-PROT112>3.3.CO;2-E;  
 RA Ghafouri B., Stahlbom B., Tagesson C., Lindahl M.;  
 RT "Newly identified proteins in human nasal lavage fluid from non-  
 RT smokers and smokers using two-dimensional gel electrophoresis and  
 RT peptide mass fingerprinting.";  
 RL Proteomics 2:112-120(2002).  
 RN [8]  
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RP MEDLINE=93393197; PubMed=7664082;  
 RA Umland T.C., Swaminathan S., Singh G., Warty V., Furey W.,  
 RA Fletcher J., Sax M.;  
 RT "Structure of a human Clara cell phospholipid-binding protein-ligand  
 RT complex at 1.9-A resolution.";  
 RL Nat. Struct. Biol. 1:538-545(1994).  
 CC -I- FUNCTION: Binds phosphatidylcholine, phosphatidylinositol,  
 CC polychlorinated biphenyls (PCB) and weakly progesterone, potent  
 CC inhibitor of phospholipase A2.  
 CC -I- SUBUNIT: Homodimer; antiparallel disulfide-linked.  
 CC -I- TISSUE SPECIFICITY: Clara cells (nonciliated cells of the surface  
 CC epithelium of the pulmonary airways).  
 CC -I- SIMILARITY: Belongs to the uteroglobin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13197; CAA31584.1; -;  
 DR EMBL; U01101; AAA81885.1; -;  
 DR EMBL; U01102; AAA18297.1; -;  
 DR EMBL; BC004481; AAH04481.1; -;  
 DR EMBL; X59875; CAA42532.1; -;  
 DR FIR; JS0036; JS0036.  
 DR FIR; S26651; S26651.  
 DR HSSP; P02779; 1UTG.  
 DR Genew; HGNC:12523; SCGB1A1.  
 DR H-InVDB; HIX0009708; -;  
 DR MIM; 192020; -;  
 DR GO; GO:0007566; P:embryo implantation; TAS.  
 DR InterPro; IPR006039; Uterogl.  
 DR InterPro; IPR003628; Uteroglbn\_sub.  
 DR InterPro; IPR000329; Uteroglbn\_subf.  
 DR InterPro; IPR006038; Uteroglbn\_subf.  
 DR Pfam; PF01099; Uteroglobin; 1.  
 DR PRINTS; PR00486; UTEROGLBIN.  
 DR ProDom; PD019935; Fel DI allergen; 1.  
 DR ProDom; PD012475; Uteroglbn\_sub; 1.  
 DR SMART; SM00096; UTG; 1.  
 DR PROSITE; PS00403; UTEROGLBIN 1; 1.  
 DR PROSITE; PS00404; UTEROGLBIN 2; 1.  
 KW Direct protein sequencing; Phospholipase A2 inhibitor; Polymorphism;  
 KW SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 91 Clara cell phospholipid-binding protein.  
 FT DISULFID 24 24 Interchain (with C-90).  
 FT DISULFID 90 90 Interchain (with C-24).  
 FT VARIANT 56 56 R -> G (in dbSNP:1802634).  
 FT VARIANT 68 68 /FTIG=VAR 012045.  
 FT VARIANT 68 68 T -> A (in dbSNP:1802632).  
 FT CONFLICT 24 24 C -> E (in Ref. 7).  
 SQ SEQUENCE 91 AA; 9994 MW; FB65ACA678F12ABD CRC64;  
 Query Match 24.4%; Score 110; DB 1; Length 91;  
 Best Local Similarity 31.0%; Pred. No. 0.00013;  
 Matches 26; Conservative 17; Mismatches 41; Indels 0; Gaps 0;  
 QY 1 MKLSVCLLLVTLALCCYQANAEFCFALVSELDFFIFSEPLFKLSLAKTDAPEVAVKL 60  
 DB 1 MKLAVTLTLVTLALCCSSASABICPSFORVIETLLMDTPSYEAAMELFSPQDMREAGA 60  
 QY 61 GVKRCTDQMSLQKRSIAEVLVKI 84  
 DB 61 QLKKLVDTLPQKPRRSIIKMEKI 84  
 RESULT 16  
 ID UTER\_RABIT STANDARD; PRT; 91 AA.  
 AC P02779;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Uteroglobin precursor (Blaetokinin).  
 GN Name=SCGB1A1; Synonyms=UGB, UGL;  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83290960; PubMed=6309802;  
 RA Bailey A., Atgier M., Atger P., Cerbon M.-A., Alison M., Vu Hai M.T.,  
 RA Logeat F., Milgrom E.;



RT "The rabbit uteroglobin gene. Structure and interaction with the  
RL progesterone receptor.";  
RJ J. Biol. Chem. 258:10384-10389(1983).  
RN (2)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83220783; PubMed=6304644;  
RA Suske G., Wenz M., Cato A.C.B., Beato M.;  
RJ "The uteroglobin gene region: hormonal regulation, repetitive elements  
RL and complete nucleotide sequence of the gene.";  
RN Nucleic Acids Res. 11:2257-2271(1983).  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83014990; PubMed=6956897;  
RA Menne C., Suske G., Arnemann J., Wenz M., Cato A.C.B., Beato M.;  
RJ "Isolation and structure of the gene for the progesterone-inducible  
RL protein uteroglobin.";  
RN Proc. Natl. Acad. Sci. U.S.A. 79:4853-4857(1982).  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83157105; PubMed=6299663;  
RA Chandra T., Bullock D.W., Woo S.L.C.;  
RJ "Hormonally regulated mammalian gene expression: steady-state level  
RL and nucleotide sequence of rabbit uteroglobin mRNA.";  
RN DNA 1:19-26(1981).  
RP SEQUENCE FROM N.A.  
RX MEDLINE=82275176; PubMed=6287481;  
RA Suske G., Menne C., Cato A., Wenz M., Beato M.;  
RJ "Characterization and sequence analysis of interspersed repetitive DNA  
RL sequences transcribed in X. laevis embryos.";  
RN Prog. Clin. Biol. Res. 85:139-146(1982).  
RP SEQUENCE OF 1-73.  
RX MEDLINE=79187160; PubMed=5717119;  
RA Atger M., Mercier J.-C., Haze G., Fridlansky F., Milgrom E.;  
RJ "N-terminal sequences of uteroglobin and its precursor.";  
RL Biochem. J. 177:985-988(1979).  
RP SEQUENCE OF 22-91.  
RX MEDLINE=79042086; PubMed=568483;  
RA Ponstingl H., Nieto A., Beato M.;  
RJ "Amino acid sequence of progesterone-induced rabbit uteroglobin.";  
RL Biochemistry 17:3908-3912(1978).  
RP SEQUENCE OF 22-91.  
RX MEDLINE=79074850; PubMed=281700;  
RA Popp R.A., Foresman K.R., Wise L.D., Daniel J.C. Jr.;  
RJ "Amino acid sequence of a progesterone-binding protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 75:5516-5519(1978).  
RP REVISIONS TO 50-62 AND 67-71.  
RA Popp R.A., Foresman K.R., Wise L.D., Daniel J.C. Jr.;  
RJ Submitted (OCT-1982) to the PIR data bank.  
RP SEQUENCE OF 22-91 FROM N.A.  
RX MEDLINE=86056319; PubMed=2415398; DOI=10.1016/0014-5793(85)80162-9;  
RA de Haro M.S., Nieto A.;  
RJ "Primary structure of rabbit lung uteroglobin as deduced from the  
RL nucleotide sequence of a cDNA.";  
RN FEBS Lett. 193:247-249(1985).  
RP SEQUENCE OF 39-77 FROM N.A.  
RX MEDLINE=81021016; PubMed=7417250;  
RA Chandra T., Woo S.L.C., Bullock D.W.;  
RJ "Cloning of the rabbit uteroglobin structural gene.";  
RL Biochem. Biophys. Res. Commun. 95:197-204(1980).  
RP SEQUENCE OF 53-72 FROM N.A.  
RX MEDLINE=80241888; PubMed=6156676;  
RA Atger M., Perricaudet M., Tiollais P., Milgrom E.;  
RJ "Bacterial cloning of the rabbit uteroglobin structural gene.";  
RL Biochem. Biophys. Res. Commun. 93:1082-1088(1980).  
RN (13)

RP X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS).  
RX MEDLINE=89199637; PubMed=2704039;  
RA Bally R., Delettre J.;  
RJ "Structure and refinement of the oxidized P21 form of uteroglobin at  
RL 1.64-A resolution.";  
RN J. Mol. Biol. 206:153-170(1989).  
RP X-RAY CRYSTALLOGRAPHY (1.34 ANGSTROMS).  
RX MEDLINE=88011213; PubMed=3656405;  
RA Morize I., Surcouf E., Vaney M.C., Epelboin Y., Buehner M.,  
RJ Fridlansky F., Milgrom E., Mornon J.-P.;  
RL "Refinement of the C222(1) crystal form of oxidized uteroglobin at  
RN 1.34-A resolution.";  
RN J. Mol. Biol. 194:725-739(1987).  
RP STRUCTURE BY NMR OF 39-68.  
RX MEDLINE=94297152; PubMed=8025221;  
RA Improtta S., Pastore A., Mammì S., Peggion E.;  
RJ "Conformation and molecular dynamics calculations on uteroglobin  
RL fragment 18-47.";  
RN Biopolymers 34:773-782(1994).  
CC -1- FUNCTION: Uteroglobin binds progesterone specifically and with  
CC high affinity. It may regulate progesterone concentrations  
CC reaching the blastocyst. It is also a potent inhibitor of  
CC phospholipase A2.  
CC -1- SUBUNIT: Homodimer; antiparallel disulfide-linked.  
CC -1- INDUCTION: Uteroglobin, synthesized in the uterus and lung, is  
CC secreted by the uterus upon induction by progesterone.  
CC -1- SIMILARITY: Belongs to the uteroglobin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; K01657; AAA31497.1; -  
CC EMBL; J00689; AAA31495.1; -  
CC EMBL; J00688; AAA31495.1; JOINED.  
CC EMBL; X01423; CAA25669.1; -  
CC EMBL; M32012; AAA31500.1; -  
CC EMBL; M25090; AAA31500.1; JOINED.  
CC EMBL; M27564; AAA31496.1; -  
CC EMBL; M25057; AAA31498.1; -  
CC EMBL; M25038; AAA31499.1; -  
CC PIR; A92391; UGRB.  
CC PDB; 1UTG; X-ray; @=22-91.  
CC PDB; 2UTG; X-ray; A/B=22-91.  
CC InterPro; IPR006039; Utergl.  
CC InterPro; IPR003628; Uteroglbn sub.  
CC InterPro; IPR000329; Uteroglbn subf.  
CC InterPro; IPR006038; Uteroglbn subf.  
CC Pfam; PF01099; Uteroglbn; 1.  
CC PRINTS; PR00486; UTEROGLBIN.  
CC ProDom; PD019935; Fel.D1 allergen; 1.  
CC ProDom; PD012475; Uteroglbn\_sub; 1.  
CC SMART; SM00096; UTG; 1.  
CC PROSITE; PS00403; UTEROGLBIN 1; 1.  
CC PROSITE; PS00404; UTEROGLBIN 2; 1.  
CC 3D-structure; Direct protein sequencing; Phospholipase A2 inhibitor;  
KW Signal; Steroid-binding.  
FT SIGNAL 1 21  
FT CHAIN 22 91 Uteroglobin.  
FT DISULFID 24 24 Interchain (with C-90).  
FT DISULFID 90 90 Interchain (with C-24).  
FT CONFLICT 6 6 T -> F (in Ref. 6).  
FT CONFLICT 16 16 C -> G (in Ref. 6).  
FT CONFLICT 46 46 L -> V (in Ref. 5).  
FT CONFLICT 67 68 DS -> NT (in Ref. 12).  
FT CONFLICT 82 82 E -> Q (in Ref. 7).  
FT HELIX 25 36



Db 1 MKIAITITVWMLSIICSSASSDICPGFLQVLEALLMESESGYVASLKPFNPGSDLQNAQT 60  
Qy 61 GVKRCTOMSLQKRSLIAEVLVKIL 85  
Db 61 QLKELVDTLPOETRINIMKLEKIL 85

## RESULT 18

UTER\_LEPCA STANDARD; PRT; 91 AA.  
ID UTER\_LEPCA  
AC P06913;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Uteroglobin precursor (Blaetokinin).  
GN Name=SCGB1A1; Synonyms=UGB, UGL;  
OS Lepus capensis (Brown hare).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Lepus.  
OX NCBT\_TaxID=9981;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=86323069; PubMed=3019311;  
RA Lopez de Haro M.S., Nieto A.;  
RT "Nucleotide and derived amino acid sequences of a cDNA coding for pre-  
uteroglobin from the lung of the hare (Lepus capensis).";  
RL Biochem. J. 235:895-898(1986).  
CC -!- FUNCTION: Uteroglobin binds progesterone specifically and with  
high affinity. It may regulate progesterone concentrations  
reaching the blastocyst. It is also a potent inhibitor of  
phospholipase A2.  
CC -!- SUBUNIT: Homodimer; antiparallel disulfide-linked.  
CC -!- INDUCTION: Uteroglobin, synthesized in the uterus and lung, is  
secreted by the uterus upon induction by progesterone.  
CC -!- SIMILARITY: Belongs to the uteroglobin family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M25609; AAA30960.1; -;  
PIR; A23825; UGBRL.  
DR HSSP; P02779; IUTG.  
DR InterPro; IPR006039; Utergl.  
DR InterPro; IPR003628; Uteroglbn sub.  
DR InterPro; IPR000329; Uteroglbn subf.  
DR InterPro; IPR006038; Uteroglbn\_supf.  
DR Pfam; PF01099; Uteroglbn; 1.  
DR PRINTS; PR00486; UTEROGLBIN.  
DR ProDom; PD01935; Fel.D1.allergen; 1.  
DR ProDom; PD012475; Uteroglbn sub; 1.  
DR SMART; SM00096; UTG; 1.  
DR PROSITE; PS00403; UTEROGLBIN\_1; 1.  
DR PROSITE; PS00404; UTEROGLBIN\_2; 1.  
KW Phospholipase A2 inhibitor; Signal; Steroid-binding.  
FT SIGNAL 1 21  
FT CHAIN 22 91 Uteroglbin.  
FT DISULFID 24 24 Interchain (with C-90).  
FT DISULFID 90 90 Interchain (with C-24).  
SQ SEQUENCE 91 AA; 9879 MW; 587614DAE9E4820F CRC64;

Query Match 22.2%; Score 100; DB 1; Length 91;

Best Local Similarity 30.2%; Pred. No. 0.0017;  
Matches 26; Conservative 16; Mismatches 44; Indels 0; Gaps 0;

Qy 1 MKLSVCLLVTLALCCQANAFPCALVSELLDFFIFSEPLFKLSLAKFDAPPAVAKL 60  
Db 1 MKLITLALVTALLCSPASAGICGFAHVNIENLLGTTPSSVETSLKEFQPDADKADGM 60

Qy 61 GVKRCTOMSLQKRSLIAEVLVKIL 86  
Db 61 QMKKVLDTLPQTRENIILKLEKIVK 86

## RESULT 19

UTER\_RAT STANDARD; PRT; 96 AA.  
ID UTER\_RAT  
AC P17559;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Clara cell phospholipid-binding protein precursor (CCBP) (Clara cells  
10 kDa secretory protein) (CC10) (Uteroglobin) (PCB-binding protein).  
GN Name=SCGB1A1; Synonyms=CC10, Ugb, Utg;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBT\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Katyal S.L., Singh G., Brown W.E., Kennedy A.L., Squeglia N.,  
Wong-Chong M.-L.;  
RT "Clara cell secretory (10 kDa) protein: amino acid and cDNA  
nucleotide sequences and developmental expression.";  
RL Prog. Respir. Res. 25:29-35(1990).  
CC [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90324266; PubMed=2115524;  
RA Nordlund-Moeller L., Andersson O., Ahlgren R., Schilling J.,  
Gallner M., Gustafsson J.-A., Lund J.;  
RT "Cloning, structure, and expression of a rat binding protein for  
polychlorinated biphenyls. Homology to the hormonally regulated  
progesterone-binding protein uteroglobin.";  
RL J. Biol. Chem. 265:12690-12693(1990).  
RN [3]  
RP SEQUENCE OF 1-18 FROM N.A.  
RX MEDLINE=90272398; PubMed=2349092;  
RA Hagen G., Wolf M., Katyal S.L., Singh G., Beato M., Suske G.;  
RT "Tissue-specific expression, hormonal regulation and 5'-flanking gene  
region of the rat Clara cell 10 kDa protein: comparison to rabbit  
uteroglobin.";  
RL Nucleic Acids Res. 18:2939-2946(1990).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
RX MEDLINE=92219263; PubMed=1560460;  
RA Umland T.C., Swaminathan S., Furey W., Singh G., Pletcher J., Sax M.;  
RT "Refined structure of rat Clara cell 17 kDa protein at 3.0-A  
resolution.";  
RL J. Mol. Biol. 224:441-448(1992).  
RN [5]  
RP STRUCTURE BY NMR.  
RX MEDLINE=96069785; PubMed=7583672;  
RA Haerd T., Barnes H.J., Larsson C., Gustafsson J.-A., Lund J.;  
RT "Solution structure of a mammalian PCB-binding protein in complex with  
a PCB.";  
RL Nat. Struct. Biol. 2:983-989(1995).  
CC -!- FUNCTION: Binds phosphatidylcholine, phosphatidylinositol,  
polychlorinated biphenyls (PCB) and weakly progesterone, potent  
inhibitor of phospholipase A2.  
CC -!- SUBUNIT: Homodimer; antiparallel disulfide-linked.  
CC -!- TISSUE SPECIFICITY: Clara cells (nonciliated cells of the surface  
epithelium of the pulmonary airways).  
CC -!- INDUCTION: By glucocorticoids.  
CC -!- SIMILARITY: Belongs to the uteroglobin family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

or send an email to license@isb-sib.ch).

CC EMBL; J05536; AAA41817.1; -;  
CC EMBL; X51318; CAA35701.1; -;  
DR PIR; A36581; A36581.  
DR PDB; 1CCD; X-ray; @=20-96.  
DR PDB; 1UTR; NMR; A/B=1-96.  
DR RGD; 3934; Scgblai.  
DR InterPro; IPR006039; Utergl.  
DR InterPro; IPR003628; Uteroglobin subf.  
DR InterPro; IPR000329; Uteroglobin subf.  
DR InterPro; IPR006038; Uteroglobin\_supf.  
DR Pfam; PF01039; Uteroglobin; 1.  
DR PRINTS; PD0486; UTEROGLBIN.  
DR ProDom; PD012475; Uteroglobin\_sub; 1.  
DR SMART; SM00096; UTG; 1.  
DR PROSITE; PS00403; UTEROGLBIN 1; 1.  
DR PROSITE; PS00404; UTEROGLBIN2; 1.  
KW 3D-structure; Phospholipase A2 inhibitor; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 96 Clara cell phospholipid-binding protein.  
FT DISULFID 24 24 Interchain (with C-90).  
FT DISULFID 90 90 Interchain (with C-90).  
FT TURN 21 22  
FT HELIX 25 34  
FT TURN 35 36  
FT HELIX 39 46  
FT TURN 47 49  
FT HELIX 53 68  
FT HELIX 71 84  
FT TURN 85 86  
FT TURN 88 90  
SQ SEQUENCE 96 AA; 10449 MW; 1A12988677B9EBEF CRC64;  
Query Match 22.2%; Score 100; DB 1; Length 96;  
Best Local Similarity 25.9%; Pred. No. 0.0016;  
Matches 22; Conservative 22; Mismatches 41; Indels 0; Gaps 0;  
QY 1 MKLSVCLLLVTLACCYQANAEFCPALVSELDFFITSEPLFKLSLAKFDAPPEAAKL 60  
DB 1 MKIAITITVLSICSSASSDIPCGLQVLEALLGSESNEYEAALKPFNPASDLQNACT 60  
QY 61 GVRCCTDMSLQKRSIAEVLVKIL 85  
DB 61 QLKRLVDLTLPQETRIINIVKILTEKIL 85  
RESULT 20  
Q8TD33 PRELIMINARY; PRT; 95 AA.  
AC Q8TD33;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Ligand binding protein RYD5.  
GN Name=RYD5;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bingle C.D.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22338213; PubMed=12406855;  
RA Reynolds S.D., Reynolds P.R., Pryhuber G.S., Finder J.D., Stripp B.R.;  
RT "Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways";  
RL Am. J. Respir. Crit. Care Med. 166:1498-1509(2002).  
RN [3]  
RP SEQUENCE FROM N.A.

TISSUE=PCR rescued clones;  
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PCR rescued clones;  
RA Strausberg R.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY026938; RAK08972.1; -;  
DR EMBL; BK002021; DAA00358.1; -;  
DR EMBL; BC069287; DAA69287.1; -;  
DR HSSP; P17559; 1CCD.  
DR InterPro; IPR006039; Utergl.  
DR InterPro; IPR006038; Uteroglobin\_supf.  
DR SMART; SM00096; UTG; 1.  
SQ SEQUENCE 95 AA; 10457 MW; F629AF06C96D2392 CRC64;  
Query Match 20.2%; Score 91; DB 2; Length 95;  
Best Local Similarity 29.5%; Pred. No. 0.019;  
Matches 28; Conservative 16; Mismatches 35; Indels 16; Gaps 2;  
QY 1 MKLSVCLLLVTLACCYQANAEFCPALVSELDFFITSEPLFKLSLAKFD 50  
DB 1 MKGRSALLLVLTLCFC-----ICRMATGEDNDFPMDFLQTLVGTPEELGKYN 54  
QY 51 APPEAAKLGVRCTDMSLQKRSIAEVLVKIL 85  
DB 55 VNEDAKAAMTELKSCRDGLQPMHKAELVKLVQVL 89  
RESULT 21  
Q8MKG2 PRELIMINARY; PRT; 91 AA.  
AC Q8MKG2;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Uteroglobin precursor.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22015297; PubMed=12021053;  
RA Muller-Schottle F., Bogusz A., Grotzinger J., Herrler A.,  
RA Krusche C.A., Beier-Hellwig K., Beier H.M.;  
RT "Full-length complementary DNA and the derived amino acid sequence of horse uteroglobin.";  
RL Biol. Reprod. 66:1723-1728(2002).  
DR EMBL; AF372660; RAM21316.1; -;  
DR HSSP; P02779; IUTG.  
DR GO; GO:0005496; F:steroid binding; IEA.  
DR InterPro; IPR006178; Fel\_DI\_allergen.

ID	UGR1_MOUSE	STANDARD;	PRT;	139 AA.
AC	Q920H1; Q920H2; Q920H3;			
AC	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Uteroglobin-related protein 1 precursor (Secretoglobin family 3A member 2) (pneumo secretory protein 1) (Pnsp-1).			
DE	Name=Scgb3a2; Synonyms=Pnsp1, UGRP1;			
GN	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS A; B AND C).			
RC	TISSUE=Lung;			
RC	MEDLINE=21539178; PubMed=11682631; DOI=10.1210/me.15.11.2021;			
RX	Niimi T., Keck-Waggoner C.L., Popescu N.C., Zhou Y., Levitt R.C.,			
RA	Kimura S.;			
RA	"UGRP1, a uteroglobin/clara cell secretory protein-related protein, is			
RT	a novel lung-enriched downstream target gene for the T/EBP/NKX2.1			
RT	homeodomain transcription factor.";			
RL	Mol. Endocrinol. 15:2021-2036(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM A).			
RC	STRAIN=NMRI;			
RA	Clippe A., Laing I.A., LeSouef P.N., Bernard A., Knoops B.;			
RA	"Molecular cloning of Pnsp-1, a protein of the respiratory tract with			
RT	potential association to atopy.";			
RT	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBUNIT: Homodimer.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=3;			
CC	Name=C;			
CC	IsolId=Q920H1-1; Sequence=Displayed;			
CC	Name=A;			
CC	IsolId=Q920H1-2; Sequence=VSP_006727, VSP_006728;			
CC	Name=B;			
CC	IsolId=Q920H1-3; Sequence=VSP_006726;			
CC	-1- TISSUE SPECIFICITY: Highly expressed in lung.			
CC	-1- SIMILARITY: Belongs to the uteroglobin family. UGRP subfamily.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).			
CC	EMBL; AF274959; AAL25708.1; -			
DR	EMBL; AF274960; AAL25709.1; -			
DR	EMBL; AF274961; AAL25710.1; -			
DR	EMBL; AF439546; AAQ04561.1; -			
DR	MGD; MGI:2153470; Scgb3a2.			
DR	GO; GO:0005576; C:extracellular; IDA.			
DR	GO; GO:0005515; F:protein binding; IPI.			
DR	InterPro; IPR006038; Uteroglobin_supf.			
DR	Pfam; PF01099; Uteroglobin; 1.			
KW	Alternative splicing; Signal.			
FT	SIGNAL 1 21 Potential.			
FT	CHAIN 22 139 Uteroglobin-related protein 1.			
FT	FT VARSPLIC 107 139 VSVLFLPMICAYPRDSKQTFATIERVFEQSKL -> EALS			
FT	FT HLK (in isoform B).			
FT	/FTId=VSP_006726.			
FT	VIITICSY -> EALSHLV (in isoform A).			
FT	/FTId=VSP_006727.			
FT	Missing (in isoform A).			
FT	/FTId=VSP_006728.			
FT	SEQUENCE 139 AA; 15431 MW; 8A2FE080B41E65E4 CRC64;			
SQL	Query Match 17.9%; Score 80.5; DB 1; Length 139;			
	Best Local Similarity 27.6%; Pred. No. 0.44;			

